DOCUMENT RESUME

ED 482 974 TM 035 344

AUTHOR	Brown, Janis
TITLE	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results for Delaware. Working Paper Series.
INSTITUTION	National Center for Education Statistics (ED), Washington, DC.
REPORT NO	No-2003-19
PUB DATE	2003-08-00
NOTE	103p.
AVAILABLE FROM	ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll Free). For full text: http://www.ed.gov/offices/OSERS/OSEP.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC05 Plus Postage.
DESCRIPTORS	Evaluation; *Reading
IDENTIFIERS	Delaware

ABSTRACT

In March 2003, the National Center for Education Statistics asked the Human Resources Research Organization to participate in a special study of 2002 reading assessment results for Delaware. This working paper summarizes the findings of that study. (AMT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 482 974

NATIONAL CENTER FOR EDUCATION STATISTICS

Working Paper Series

NAEP QUALITY ASSURANCE CHECKS OF THE 2002 READING ASSESSMENT RESULTS FOR DELAWARE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Working Paper No. 2003-19 August 2003

Contact: Janis Brown
Assessment Development and Quality Assurance
Janis.Brown@ed.gov

U. S. Department of Education
Institute of Education Sciences

TM035344

02

BEST COPY AVAILABLE

U.S. Department of Education
Rod Paige
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Val Plisko
Associate Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006–5651

August 2003

The NCES World Wide Web Home Page address is *http://nces.ed.gov*
The NCES World Wide Web Electronic Catalog is: *http://nces.ed.gov/pubsearch*

Suggested Citation

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. *NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware*. NCES 2003–19, by R. Gene Hoffman, D.E. (Sunny) Becker, and Lauress L. Wise. Project Officer: Janis Brown. Washington, DC: 2003.

For ordering information on this report, write:

U.S. Department of Education
ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

Or call toll free 1–877–4ED–Pubs

Content Contact:
Janis Brown
(202) 502–7419
Janis.Brown@ed.gov

Foreword

In addition to official NCES publications, NCES staff and individuals commissioned by NCES produce preliminary research reports that include analyses of survey results, and presentations of technical, methodological, and statistical evaluation issues.

The ***Working Paper Series*** was initiated to promote the sharing of the valuable work experience and knowledge reflected in these preliminary reports. These reports are viewed as works in progress, and have not undergone a rigorous review for consistency with NCES Statistical Standards prior to inclusion in the Working Paper Series.

Copies of Working Papers can be downloaded as pdf files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/), or contact Sheilah Jupiter at (202) 502–7363, e-mail: sheilah.jupiter@ed.gov, or mail: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 1990 K Street NW, Room 9048, Washington, DC 20006.

Marilyn M. Seastrom
Chief Mathematical Statistician
Statistical Standards Program

Ralph Lee
Mathematical Statistician
Statistical Standards Program

NAEP Quality Assurance Checks of the 2002 Reading Assessment Results for Delaware

EXECUTIVE SUMMARY

In March 2003, the National Center for Education Statistics (NCES) asked the Human Resources Research Organization (HumRRO) to participate in a special study of the 2002 reading assessment results for Delaware. Standard review of test results had revealed that compared with other states, Delaware (DE) was an outlier from the mainstream, both in the change in exclusion rates between 1998 and 2002, and in the 4th grade reading gains between 1998 and 2002, particularly for the Delaware Hispanic population. NCES authorized several teams to investigate various aspects of the assessment. HumRRO was asked to focus on seven specific technical questions, and follow any additional data analysis leads that emerged. Below is a summary of findings for each question.

Question 1: Was there a problem with the sampling of Delaware students?

We found no problems with the sampling of Delaware students. We investigated the sampling process in two ways. First, an expert sampling statistician reviewed the 2002 sampling for Delaware and concluded that there were no problems; inclusion of all Delaware schools led to increased accuracy and did not in and of itself increase or decrease score estimates. Second, the weighted count of students from the NAEP sample was closely comparable to enrollment counts from the Delaware Department of Education.

Question 2: Was there a problem with the weighting [case weights] of the Delaware data?

We detected no problem with the case weights of the Delaware data. Delaware is one of the few states where every school is sampled and in 2002 nearly all of the students in the targeted grades were tested. Consequently, the sampling weight assigned to each school should be 1.0, and they were exactly that on the 2002 data file. In addition, student weights should all be the same except for minor differences due to reassignment of the weights for students who were absent. The 2002 student weights were found to be entirely consistent with this expectation.

Question 3: Was there a problem with the design for assigning test booklets to students (BIB spiral)?

No problem with the BIB (balanced incomplete block) spiral was detected. Booklets and items were distributed appropriately across the state, as well as within each school. The distribution of booklets in Delaware schools closely matched the distribution in other states.

Question 4: Was there a problem with the scoring (hand scoring or scanning) of the Delaware data?

We found no problem in the scoring of Delaware data. Open-ended responses from Delaware students were mixed in with responses from other states in the scoring process; there was no differential treatment. Similar treatment was also found for the scanning and scoring of responses to the multiple-choice questions. Delaware students did not have unusual gains on any open-ended or multiple-choice item, which might have indicated a problem with the scoring of that item.

However, Delaware students did show slightly larger gains between 1998 and 2002 on the open-ended items relative to the rest of the nation. This difference might be due to a greater emphasis on writing, both in the instruction process and in the state's own assessment, and may account for some of the gain seen by Delaware.

Question 5: Was there an error in the scaling and equating for Delaware?

No scaling or equating problems were identified in Delaware. Several analyses examined patterns of item performance and scale scores for Delaware and the rest of the nation. The relationship between scores on the individual items to scale score estimates was the same for Delaware as for other states.

Question 6: Was there a problem with the coding of any data in Delaware?

We found no coding problems. Race/ethnicity codes used for reporting were reviewed because of large gains by Hispanics. Agreement between race/ethnicity data supplied by students and by schools was sufficient to rule out coding errors, overall and for each school.

Question 7: Was there a breach in test security in Delaware?

No indications of test security breaches were identified. Gains on individual items and on blocks of items associated with a common passage were consistent with gains on these items and blocks for the nation as a whole. Individual schools did not show unusual gains on individual items, blocks of items, or overall.

Additional Exploration: Were any other problems detected that would suggest interpreting the 1998-2002 results with caution?

Prior to calculating the gains between 1998 and 2002, the 1998 results were recomputed (1) using an alternate sample of students who were provided accommodations similar to those provided in 2002 and (2) defining race categories from codes supplied by schools rather than students. Consequences of these changes in the 1998 data were:

- Grade 4 sample size for Hispanics decreased from 198 to 101.
- The exclusion rate for Grade 4 Hispanics dropped from 6 percent to just 3 percent.
- Grade 8 sample size for Hispanic students decreased from 78 to 64.
- For Grade 8, the exclusion rate for Hispanics dropped from 12 percent to 0 percent.

The "2002 gains" were based upon these recomputed 1998 scores. Gains between 2002 scores and recomputed 1998 scores had large standard errors and therefore wide confidence bands:

- The 95 percent confidence band for Grade 4 Hispanic gain is +13 to +59 points.
- The 95 percent confidence band for Grade 8 Hispanic gain is -14 to +18 points.

CONCLUSION: Based on an extensive analysis of the 2002 Delaware reading assessment data and on data from the 1998 assessment used as the basis for computing gains in 2002, we did not find any technical/analytic problems in data sampling or analysis that affected the 2002 results for Delaware.

We did note that recomputed 1998 score estimates for 4th and 8th grade Hispanic students were based on small sample sizes, large standard errors, and low exclusion rates. Consequently, the score gains between 2002 and recomputed 1998 had wide confidence bands. We recommend that the Delaware Hispanic gains for Grade 4 and Grade 8 from 1998 to 2002 be flagged, with explanatory text, to indicate that the amount of gain should be interpreted with caution.

Table of Contents

EXECUTIVE SUMMARY i

CHAPTER 1: INTRODUCTION 9
- HumRRO Analysis Goals 10
- Methodology 11
- Foreshadowing the results 11

CHAPTER 2: SAMPLING 13
- Question 1: Was there a problem with the sampling of Delaware students? 13
- Theoretical Sampling Issues for Delaware Schools and Students in 2002 13
- Grade 4 Analyses 13
  - Examination of the Data – Accounting for Students 13
  - Examination of the Data – Accounting for Students by Subsample 14
  - Examination of the Data – Sampling and Score Means 17
- Grade 8 Analyses 19
- Delaware Sampling Conclusions 22

CHAPTER 3: WEIGHTING 25
- Question 2: Was there a problem with the weighting [case weights] of the Delaware data? 25
- Delaware Weighting Conclusions 25

CHAPTER 4: BIB SPIRAL 27
- Question 3: Was there a problem with the design for assigning test booklets to students (BIB spiral)? 27
- Delaware BIB Spiraling Conclusions 27

CHAPTER 5: SCORING 31
- Question 4: Was there a problem with the scoring (hand scoring or scanning) of the Delaware data? 31
- Delaware Scoring Conclusions 66

CHAPTER 6: SCALING AND EQUATING 67
- Question 5: Was scaling and equating performed correctly in Delaware? 67
- Delaware Scaling and Equating Conclusions 73

CHAPTER 7: CODING 75
- Question 6: Was there a problem with the coding of any data in Delaware? 75
- Delaware Coding Conclusions 77

CHAPTER 8: TEST SECURITY 79
- Question 7: Was there a breach in test security in Delaware? 79
- Delaware Security Conclusions 82

CHAPTER 9: CONCLUSIONS 85

APPENDIX A: OUTSIDE REVIEW OF NAEP SAMPLING PROCEDURES 87
- Samples versus Censuses 87
- Summary 88

List of Tables

Table 1.1. Score Gains for NAEP Reading 1998–2002 9
Table 2.1. 1998 NAEP Reading Assessment Samples 14
Table 2.2. 1998 NAEP Reading Assessment R2 Sample 15
Table 2.3. 1998 NAEP Reading Assessment R3 Sample 15
Table 2.4. Grade 4 1998 S2 and S3 Sample Sizes: Within Delaware and Outside Delaware 16
Table 2.5. Grade 4 1998 S2 and S3 Sample Sizes: Hispanics Only 17
Table 2.6. Grade 4 1998 SD/LEP Assessed Students 17
Table 2.7. Grade 4 1998 Delaware Hispanic Scale Scores Computed Under Four Conditions ... 18
Table 2.8. Grade 4 1998 Means for S2 and S3 Samples 19
Table 2.9. Grade 4 1998 Raw Proportion Correct for S2 and S3 Samples 19
Table 2.10. Grade 8 S2 and S3 Sample Sizes: Within Delaware and Outside Delaware 20
Table 2.11. Grade 8 S2 and S3 Sample Sizes: Hispanics Only 20
Table 2.12. Grade 8 SD/LEP Assessed Students 21
Table 2.13. Grade 8 1998 Delaware Hispanic Scale Scores Computed Under Four Conditions . 21
Table 2.14. Grade 8 1998 Means for S2 and S3 Samples 22
Table 2.15. Grade 8 1998 Raw Proportion Correct for S2 and S3 Samples 22
Table 2.16. Standard Errors for Delaware Hispanic Reading Gains 23
Table 2.17. Standard Errors for Delaware Reading Gains: All Students 23
Table 4.1. Distribution of 2002 4th Grade Reading Books in Delaware 28
Table 4.2. Distribution of 2002 8th Grade Reading Books in Delaware 29
Table 7.1. Comparison of Race/Ethnicity Codes from Schools and Students in the 2002 Delaware 4th Grade Reading Assessment 76
Table 7.2. Comparison of Race/Ethnicity Codes from Schools and Students in the 2002 Delaware 8th Grade Reading Assessment 76
Table 7.3. Discrepancies between School and Student Race/Ethnicity Codes for Individual Delaware Schools in the NAEP 2002 4th Grade Reading Assessment 77
Table 7.4. Discrepancies between School and Student Race/Ethnicity Codes for Individual Delaware Schools in the NAEP 2002 8th Grade Reading Assessment 77

List of Figures

Figure 5.1. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for all students 32
Figure 5.2. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Hispanic students 33
Figure 5.3. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Black students 34
Figure 5.4. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for White students 35
Figure 5.5. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for all students 36
Figure 5.6. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Hispanic students 37

Figure 5.7. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Black students 38
Figure 5.8. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for White students 39
Figure 5.9. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for all students 41
Figure 5.10. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students 42
Figure 5.11. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Black students 43
Figure 5.12. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for White students 44
Figure 5.13. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for all students 45
Figure 5.14. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students 46
Figure 5.15. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Black students 47
Figure 5.16. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for White students 48
Figure 5.17. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for all students 49
Figure 5.18. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Hispanic students. 50
Figure 5.19. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Black students 51
Figure 5.20. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for White students 52
Figure 5.21. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for all students 54
Figure 5.22. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Hispanic students 55
Figure 5.23. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Black students 56
Figure 5.24. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for White students 57
Figure 5.25. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for all students 58
Figure 5.26. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students 59
Figure 5.27. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Black students 60
Figure 5.28. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for White students 61
Figure 5.29. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for all students 62

Figure 5.30. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students 63
Figure 5.31. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Black students. 64
Figure 5.32. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for White students 65
Figure 6.1. "Test characteristic curve" captured from the 2002 Grade 4 reading data 68
Figure 6.2. "Test characteristic curve" captured from the 2002 Grade 8 reading data 69
Figure 6.3. Relationship between 2001 state reading scores for Grade 3 and 2002 NAEP reading scores for Grade 4 for Delaware schools 70
Figure 6.4. Relationship between 2001 state reading scores for Grade 5 and 2002 NAEP reading scores for Grade 4 for Delaware schools 71
Figure 6.5. Relationship between 2001 state reading scores for Grade 8 and 2002 NAEP reading scores for Grade 8 for Delaware schools 72
Figure 8.1. Delaware 1998–2002 4th grade school gains on raw score versus gain on NAEP scale scores 80
Figure 8.2. Non-Delaware 1998–2002 4th grade school gains on raw score versus gain on NAEP scale scores 81
Figure 8.3. Delaware 1998–2002 8th grade school gains on raw score versus gain on NAEP scale scores 83
Figure 8.4. Non-Delaware 1998–2002 8th grade school gains on raw scores versus gain on NAEP scale scores 84

NAEP Quality Assurance Checks of the 2002 Reading Assessment Results for Delaware

CHAPTER 1: INTRODUCTION

In March 2003, the National Center for Education Statistics (NCES) asked the Human Resources Research Organization (HumRRO) to participate in a special study of the 2002 reading assessment results for Delaware. Standard review of test results had revealed that compared with other states, Delaware was an outlier from the mainstream, both in the increase in exclusions between 1998 and 2002, and in the 4th grade reading gains between 1998 and 2002, particularly for the Delaware Hispanic population. NCES authorized several teams to investigate various aspects of the assessment. Preliminary data presented to HumRRO suggested that the gains are particularly noticeable for Grade 4 and most extreme for the Grade 4 Hispanic students. Table 1.1 presents updated results computed after the 2002 data had received its final edits. Using updated data, these results defined the issue that HumRRO was asked to address.

Table 1.1. Score Gains for NAEP Reading 1998–2002

Score gains for NAEP Reading 1998 - 2002
(Computed by HumRRO)

Grade 4					
Year	All States/ All Students	Delaware			
		All Students	White	Black	Hispanic
1998 Mean[a]	215	212	220	199	193
1998 Mean-R[b]	213	207	218	189	176
2002 Mean	219	224	233	209	212
Gain[c]	6	17	15	20	36
Grade 8					
	All States/ All Students	Delaware			
		All Students	White	Black	Hispanic
1998 Mean[a]	262	256	263	238	246
1998 Mean-R[b]	261	254	263	234	248
2002 Mean	264	268	275	252	250
Gain[c]	3	14	12	18	2

[a] 1998 Mean is computed for students who were not provided with accommodations and whose race/ethnicity was based on student-reported data.
[b] 1998 Mean-R is computed for students who were provided with accommodations and whose race/ethnicity was based on school-reported data.
[c] Gain is 2002 Mean minus 1998 Mean-R.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 1.1 presents two sets of data for 1998. That year included two separate subsamples for a study of the impact of testing with accommodations. Original reports for the 1998 assessment were based on a subsample in which students were not allowed accommodations, consistent with practices in earlier assessments. At the same time, approximately half of the

students were tested under test administration rules that allowed accommodations. Since the 2002 assessment did allow accommodations, it is the 1998 accommodated sample that is most appropriate for comparisons to 2002 achievement; therefore, results for 1998 were recomputed (denoted "Mean-R" in Table 1.1) based on the subsample for which accommodations were allowed in 1998. Furthermore, the 2002 mean is based on information about the race/ethnicity supplied by the schools rather than each student's response to background questionnaire items. The 1998 mean was initially based on student response information. The recomputed 1998 Mean-R reflected this change to school-based race/ethnicity determination.

Several observations that help frame the issue of the Delaware gain can be made from the table. First, the table reveals that recomputing 1998 scores did change the results. For example, comparison of the "1998 Mean" row to the recomputed "1998 Mean-R" row indicates that the difference for the nation as a whole is small (i.e., Grade 4 decreased by two points from 215 to 213; Grade 8 decreased by one point from 262 to 261). On the other hand, the difference between the 1998 Mean and the 1998 Mean-R is larger for Delaware, particularly for Grade 4, with a 17-point difference for Hispanics (from 193 to 176). Thus, had Hispanic gains between 1998 and 2002 been computed on the original 1998 scores, the gain would still be large (212 minus 193, or 19 points), but not as large as the 36-point gain (212 minus 176) being reported.

For Grade 8, the atypical result appears to be that Delaware Hispanics did not gain like the rest of Delaware. In addition, the difference between the 1998 Mean and the 1998 Mean-R was in the opposite direction from the differences in the rest of the table.

The definition of the potential problem posed to HumRRO had two parts:

- the difference between Delaware as a whole and the rest of the nation and,
- within Delaware, the difference between Hispanics and the other race/ethnicity categories.

In addition, the data in Table 1.1 suggest that the recomputation of scores for 1998 behaved differently for Delaware than the rest of the nation, again particularly for Hispanics. Therefore, questions about Delaware gains concern both the 2002 and 1998 assessments.

HumRRO Analysis Goals

Because of the size of the Delaware gains, NCES commissioned four teams to investigate four aspects of the assessment:

- Delaware context
- Technical issues
- Exclusions
- Options for reporting

HumRRO was assigned seven specific technical questions and was asked to follow any additional data analysis leads that emerged.

- Question 1: Was there a problem with the sampling of Delaware students?

- Question 2: Was there a problem with the weighting [case weights] of the Delaware data?
- Question 3: Was there a problem with the design for assigning test booklets to students (BIB spiral)?
- Question 4: Was there a problem with the scoring (hand scoring or scanning) of the Delaware data?
- Question 5: Was scaling and equating performed correctly in Delaware?
- Question 6: Was there a problem with the coding of any data in Delaware?
- Question 7: Was there a breach in test security in Delaware?

One chapter is devoted to each of these issues. This paper is intended to be accessible to a non-technical audience. Therefore, although some technical details are critical to the full explanation of our findings, numerous visual aids are included to help clarify the results.

Methodology

The bulk of this effort involved independent analyses of data provided by NCES, Educational Testing Service (ETS), Westat, and Pearson Educational Measurement. In some cases, efforts were made to reproduce results exactly. In other instances, targeted analyses investigated related issues to produce confirmatory/divergent evidence. Some of the analyses address more than one of the above questions. Most of the analyses were focused on Delaware Grade 4 but some parallel analyses were also conducted for Grade 8.

Foreshadowing the results

Unfortunately, HumRRO did not have all of the details of the two 1998 sampling conditions for 1998 until substantial analyses had been completed. As it turned out, the split sampling in 1998 was one of the keys for understanding Delaware gains, especially for Hispanics. On the other hand, the 1998 split sample does not appear to be the whole story, and our in-depth analysis of 2002 technical issues provides some confirmatory evidence about the Delaware gains in general.

CHAPTER 2: SAMPLING

Question 1: Was there a problem with the sampling of Delaware students?

Because Delaware is a relatively small state, all schools were included in NAEP assessment in 1998 and 2002. However, the sampling of students within schools differed across the two assessment periods. In 1998, students were sampled, with target numbers within each school set for either one or two test administration sessions (i.e., 32 students or 64 students), depending on the size of the school. In 2002, all students were eligible for testing.

Two types of questions emerge, theoretical and actual. *Theoretical* questions concern sampling theory and the extent to which sampling, per se, could systematically increase or decrease achievement level estimates. For example, did the fact that all Delaware schools participated in the 2002 assessment provide any statistical advantage over other states where only a sample of schools participated? Similarly, did the fact that all Delaware students were targeted for testing in 2002 provide any statistical advantage over sampling of Delaware students in 1998? (For a more complete explanation of sampling theory, see Appendix A.)

The second question concerns the *actual* characteristics of the tested population. This question is complicated by the split sample for 1998. In addition, when 1998 scores were recomputed, there was one other change that most directly affects the score distribution estimates for the different race/ethnicity categories. For 2002, results are based on the information about race/ethnicity supplied by the schools rather than each student's responses to background questionnaire items. Original 1998 results were based on student response information. Recomputed 1998 score distributions included this change to school-based race/ethnicity determination.

Theoretical Sampling Issues for Delaware Schools and Students in 2002

Dr. Chuck Cowan of Analytic Focus provided us an overview of theoretical issues pertaining to sampling procedures and described any differences that might be associated with testing in all schools rather than just a sample. Dr. Cowan has extensive experience, both at NCES and at the Bureau of the Census, working on thorny sampling issues and is also a consultant to the Department of Defense on sampling issues associated with the recent renorming of the Armed Services Vocational Aptitude Battery. Dr. Cowan's response, reproduced in Appendix A, indicates that testing more students in more schools would increase the overall accuracy of score estimates but would not affect estimates of average scores in any consistent way. On the other hand, it is well known that samples that are too small may provide inconsistent or unstable results, of particular concern for subgroups.

Grade 4 Analyses

Examination of the Data – Accounting for Students

HumRRO also compared estimated counts of Delaware Grade 4 students generated from the NAEP Grade 4 2002 Reading data file against the Delaware Department of Education

population statistics for Grade 5 in September 2002.[1] These Grade 5 students would have been in Grade 4 during the 2002 testing window. The Delaware Department of Education reports 9,089 students began Grade 5 this year. The NAEP 2002 reading assessment sample includes 4,185 Delaware students. Slightly more than half of the students in Delaware participated in other assessments or special studies (e.g., mathematics). Thus, one cannot simply compare the number of students in the reading assessment to the state counts. Instead, a weighted count was computed, where each student tested also represented some (slightly more than one) of the other students who participated in a different assessment. The resulting count was 8,283. Note that NAEP does not attempt to represent students who cannot be assessed. In Delaware, 9 percent of the students selected were excluded because they could not be assessed. Thus, the appropriate comparison count was 91 percent of the 9,089, or 8,271. Given transfers in and out of the state after NAEP testing and a few students who may have been retained in grade, the NAEP estimate is very close to the counts reported by Delaware. Consequently, it is reasonable to conclude that NAEP did account for essentially all of Delaware's Grade 4 students in the 2002 Reading assessment.

A school-by-school accounting for 4th grade students was conducted in which student counts from the NAEP data file (linked with school name information provided by ETS) were matched to school population counts from the Delaware Web site. This analysis provided confirmatory evidence that a census test of 4th grade students was conducted.

Examination of the Data – Accounting for Students by Subsample

By design, the 1998 NAEP Reading assessment divided all sampled students equally into two distinct samples. For historical reasons, these samples are labeled S2 and S3. (An earlier study also had a sample, labeled S1, for which a more limited inclusion policy was used. The impact of the difference in inclusion policy was found to be minimal, so this condition was dropped from the 1998 study.) Within each of the two 1998 samples, students were subdivided into three samples on the basis of information about their status with respect to disabilities (SD) and limited English proficiency (LEP) and whether they were administered the test. As a result, students are identified according to the schema in Table 2.1.

Table 2.1. 1998 NAEP Reading Assessment Samples

1998 NAEP Reading Assessment Sample Codes		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	A2	A3
SD/LEP ASSESSED	B2	B3
DS/LEP EXCLUDED	C2	C3

[a] Sample tested with accommodations not permitted

[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.1 shows the two samples – one for which accommodations were not allowed (S2) and one for which they were (S3). Within each of those samples are three subsamples of

[1] See http://www.doe.state.de.us/reporting/enrollment/0203/Unit%20count-Enrollment%20by%20grade;%20Sept.pdf.

students: (A) students who were neither SD nor LEP and consequently took the assessment in a regular session, (B) students who were either SD or LEP and could be assessed without accommodations or with the accommodations offered as dictated by their main sample designation, and (C) SD and LEP students who could not be assessed because they required accommodations not available in their main sampling condition. The S2 and S3 subsamples were determined at the school level, that is, schools were either in the S2 sample or the S3 sample, and so the accommodation policy was the same for all students within any school.

State scores for 1998 were originally reported using students from the cells labeled A2, A3, and B2 (see Table 2.2). To make comparisons to 2002 state data, 1998 data were recomputed using students from cells A2, A3, and B3 (see Table 2.3). Thus, the B2 students tested with no accommodations allowed were replaced by the B3 students who were tested with allowed accommodations, as needed. The set of students that include A2, A3, and B2 is labeled reporting sample R2. The set of students that include A2, A3, and B3 is labeled reporting sample R3.

Table 2.2. 1998 NAEP Reading Assessment R2 Sample

1998 NAEP Reading Assessment Sample Codes		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	**A2**	**A3**
SD/LEP ASSESSED	**B2**	B3
SD/LEP EXCLUDED	C2	C3

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.3. 1998 NAEP Reading Assessment R3 Sample

1998 NAEP Reading Assessment Sample Codes		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	**A2**	**A3**
SD/LEP ASSESSED	B2	**B3**
SD/LEP EXCLUDED	C2	C3

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

The following tables illustrate the effects of the accommodation policy on samples for Delaware and all states other than Delaware. (The percentages indicate percentage within the column.) The sample size data in Table 2.4 reveal that in states other than Delaware the policy did not shift the proportions between SD/LEP students who were assessed and SD/LEP students who were not assessed (i.e., excluded). On the other hand, for the S3 sample compared to the S2 for Delaware, the SD/LEP assessed rose from 9 percent to 14 percent, while the SD/LEP excluded students dropped from 8 percent to 1 percent. The ratio of SD/LEP students assessed to

those excluded in Delaware is 14 to 1, contrasting markedly not only with non-Delaware states, but also to the Delaware S2 sample.

Having only 1 percent of the Delaware S3 sample excluded seems more consistent with a policy of testing all SD/LEP students unless they are in the severely disabled population, which tends to be 1–2 percent of the general population. Certainly, this change in exclusion rate raises a question about the test exclusion practices implemented in Delaware's S3 schools versus S2 schools. It also raises a concern about comparing Delaware's exclusion rates in 2002 to those in 1998, because two distinct sets of rates appear to have been operating in 1998.

Table 2.4. Grade 4 1998 S2 and S3 Sample Sizes: Within Delaware and Outside Delaware

	Non-Delaware Sample Sizes	
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	52,965 (85%)	52,505 (85%)
SD/LEP ASSESSED	4,359 (7%)	4,937 (8%)
EXCLUDED	5,094 (8%)	4,029 (6%)
	Delaware Sample Sizes	
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	1,099 (83%)	1,086 (85%)
SD/LEP ASSESSED	124 (9%)	174 (14%)
EXCLUDED	109 (8%)	18 (1%)

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.5 examines sample sizes for Hispanics. Because NCES has shifted to defining race by school report as a more accurate indicator, only school-reported race is included. Similar to the above tables, the proportion of SD/LEP and excluded Hispanic students are roughly equal for the S2 and S3 samples for states other than Delaware but are markedly different within Delaware. Following the pattern of the state as a whole, few Hispanic students in the S3 sample were excluded. Furthermore, the proportion of non-SD/LEP students was lower in the S3 sample than in the S2 sample. The table also shows that the actual Hispanic sample sizes are small.

Table 2.6 includes some test administration information about the SD/LEP assessed Hispanic students. For both Delaware and non-Delaware states, the data indicate that only about 18 percent of the SD/LEP tested students actually received accommodations, that is, the shift from the S2 to the S3 sample is operative for a relatively small number of students. In Delaware's S3 sample, only five students took the Reading assessment with an accommodation.

Table 2.5. Grade 4 1998 S2 and S3 Sample Sizes: Hispanics Only

Non-Delaware Hispanic Sample Sizes		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	3,349 (60%)	3,649 (62%)
SD/LEP ASSESSED	1,054 (19%)	1,091 (19%)
EXCLUDED	1,174 (21%)	1,128 (19%)
Delaware Hispanic Sample Sizes		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	34 (72%)	39 (57%)
SD/LEP ASSESSED	6 (12%)	28 (41%)
EXCLUDED	7 (15%)	2 (3%)

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.6. Grade 4 1998 SD/LEP Assessed Hispanic Students

SD/LEP Assessed Students – Test Administration		
	S2 Sample[a]	S3 Sample[b]
Non-Delaware Hispanic	Total = 1,054 98.8% in regular session. 1.2% in makeup session None with accommodations	Total = 1,091 82.4% in regular session, without accommodations 1.1% in regular makeup session 9.4% large print 6.0% small group 1.0% other accommodations
Delaware Hispanic	Total = 6 100% in regular session	Total = 28 82.1% in regular session without accommodations 14.3% (4 students all IEP students) large print 3.6% (1 LEP student) in small group

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Clearly, there is a concern here about the meaning of the sampling change when it involves so few students. On the other hand, perhaps the change has little or no meaning for estimating Delaware scores. The following tables examine test performance data.

Examination of the Data – Sampling and Score Means

The impact of the changes in testing accommodations and ethnic determination for Delaware Hispanics is evidenced in Table 2.7. The effects of the change in racial/ethnic

determination on average scores are inconsistent. On the other hand, allowing accommodations appears to reduce scores whichever way race is determined. Using school-reported race, there is a 26-point difference between the accommodated (S3) and non-accommodated (S2) samples. Since the purported Delaware Hispanic gain for 1998 to 2002 is 36 scale points, a large part of that gain may be related to the accommodation testing conditions in Delaware.

Table 2.7. Grade 4 1998 Delaware Hispanic Scale Scores Computed Under Four Conditions

	Self-Reported Race/Ethnicity	School-Reported Race/Ethnicity
No accommodations allowed	193 (standard error = 3.8) (n=198) *(Original Score, R2 Sample)*	202 (standard error = 5.5) (n=79)
Accommodations allowed	184 (standard error = 7.5) (n=184)	176 (standard error = 11.6) (n=101) *(Recomputed Score, R3 Sample)*

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.8 shows unweighted performance means in order to more directly examine the students who actually took the test. Two different mean scores are presented—one based on the R2 sample (the A2, A3, and B2 students, labeled *Original*) and one based on the R3 sample (the A2, A3, and B3 students, labeled *Recomputed*). For states other than Delaware, mean plausible values changed based on the sample change, but only by 1 scale-score point. On the other hand, larger changes are apparent for Delaware, particularly for SD/LEP assessed students. The shift in testing with accommodations had a much larger impact in Delaware than in the rest of the nation as a whole. In fact, the shift in accommodation policy (which again was operative for only 5 Hispanic students), appears to reduce the SD/LEP portion of the Hispanic sample mean by 43 scale points. If this change is coupled with the large change in Hispanic exclusion rates for the S3 sample, the data suggest that the exclusion rule applied to the Delaware S2 sample was not applied to the Delaware S3 sample. It appears as if students who would have been excluded in the S2 sample were tested in the S3 sample. As a result, the means for tested SD/LEP dropped noticeably. Obviously, this would have an impact on the apparent gains for Hispanic students between 1998 and 2002.

For verification of the plausible value results, test performance was also examined using student raw item response data (see Table 2.9). Since students take test forms of different lengths and with different mixes of multiple-choice and open-ended items, maximum possible points are not constant across students. Therefore, raw score performance was calculated as the proportion of total possible points that a student earned, where total points varied by test form. Again, the difference between testing with accommodations allowed versus without accommodations allowed was small in states other than Delaware. In Delaware the difference is more apparent.

Table 2.8. Grade 4 1998 Means for S2 and S3 Samples

Non-Delaware Hispanic Plausible Value Scale Score Means – Unweighted (Original/Recomputed)		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	204/203[c]	
SD/LEP ASSESSED	171/--	--/170
EXCLUDED	No score on file	
Delaware Hispanic Plausible Value Scale Score Means – Unweighted (Original/Recomputed)		
	S2 Sample	S3 Sample
Non SD/LEP	204/200[b]	
SD/LEP ASSESSED	193/--	--/150
EXCLUDED	No score on file	

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted
[c] Because the non-SD/LEP students were used in both reporting samples, the original and recomputed means for these students include students from both the S2 and S3 samples.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.9. Grade 4 1998 Raw Proportion Correct for S2 and S3 Samples

Non-Delaware Hispanic Raw Proportion Correct		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	.39	
SD/LEP ASSESSED	.243	.242
EXCLUDED	.0003	
Delaware Hispanic Raw Proportion Correct		
	S2 Sample[a]	S3 Sample[b]
Non SD/LEP	.37	
SD/LEP ASSESSED	.300	.165
EXCLUDED	0	

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Grade 8 Analyses

Tables 2.10 and 2.11 present comparable information for Grade 8 students. These tables show that, similar to Grade 4, exclusion rates in Delaware appear inconsistent for the S3 sample of the Delaware students overall and for Hispanic students. For Delaware Grade 8 Hispanics, a large change is also seen in the proportion of students in the non-SD/LEP group—from 57 percent to 88 percent. Like Grade 4, it is apparent that the actual number of tested Hispanic

students in Delaware is small. None of the Grade 8 Hispanic students were excluded from testing.

Table 2.10. Grade 8 S2 and S3 Sample Sizes: Within Delaware and Outside Delaware

	Non-Delaware Sample Sizes	
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	44,088 (87%)	44,824 (87%)
SD/LEP ASSESSED	3,530 (7%)	4253 (8%)
EXCLUDED	3,349 (7%)	2640 (5%)
	Delaware Sample Sizes	
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	952 (82%)	930 (89%)
SD/LEP ASSESSED	105 (9%)	94 (9%)
EXCLUDED	102 (9%)	20 (2%)

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.11. Grade 8 S2 and S3 Sample Sizes: Hispanics Only

	Non-Delaware Hispanic Sample Sizes	
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	3,219 (70%)	3,243 (71%)
SD/LEP ASSESSED	742 (16%)	785 (17%)
EXCLUDED	668 (14%)	569 (12%)
	Delaware Hispanic Sample Sizes	
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	39 (57%)	22 (88%)
SD/LEP ASSESSED	18 (26%)	3 (12%)
EXCLUDED	11 (16%)	0 (0%)

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.12 shows that, as with Grade 4, only about 79 percent of the SD/LEP Hispanic students in states other than Delaware took the assessment with accommodations. However, there were only 3 SD/LEP Hispanics in Grade 8 in the Delaware S3 sample, and none of them used an accommodation. Clearly, one is dealing with very small numbers when considering the effects of the accommodation policy on Delaware Hispanics.

Table 2.12. Grade 8 SD/LEP Assessed Students

	SD/LEP Assessed Students – Test Administration	
	S2 Sample[a]	S3 Sample[b]
Non-Delaware Hispanic	Total = 742 97.1% in regular session 2.8% in regular makeup session None with accommodations	Total = 785 79.4% in regular session 1.9% in regular makeup session 11.6% large print 5.7% small group 1.4% other accommodations
Delaware Hispanic	Total = 57 94.4% in regular session 5.6% in regular makeup session	Total = 3 100% in regular session

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Test results for Grade 8 (Tables 2.13 through 2.15) appear more stable across sampling conditions than they did for Grade 4 and show less change based on how race/ethnicity was determined.

Table 2.13. Grade 8 1998 Delaware Hispanic Scale Scores Computed Under Four Conditions

	Self-Reported Race/Ethnicity	School-Reported Race/Ethnicity
No accommodations allowed	246 (standard error = 8.7) (n = 78) *(Original Scores, R2 Sample)*	247 (standard error = 8.6) (n = 79)
Accommodations allowed	247 (standard error = 8.2) (n = 63)	248 (standard error = 7.9) (n=64) *(Recomputed Scores, R3 Sample)*

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.14. Grade 8 1998 Means for S2 and S3 Samples

Non-Delaware Hispanic Plausible Value Scale Score Means – Unweighted (Original/Recomputed)		
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	254/254 [c]	
SD/LEP ASSESSED	221	221
EXCLUDED	No score	
Delaware Hispanic Plausible Value Scale Score Means – Unweighted (Original/Recomputed)		
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	257/257 [c]	
SD/LEP ASSESSED	217	218
EXCLUDED	No score	

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted
[c] Because the non-SD/LEP students were used in both reporting samples, the original and recomputed means for these students include students from both the S2 and S3 samples.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 2.15. Grade 8 1998 Raw Proportion Correct for S2 and S3 Samples

Non-Delaware Hispanic Raw Proportion Correct		
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	0.48 [c]	
SD/LEP ASSESSED	0.31	0.32
EXCLUDED	0	
Delaware Hispanic Raw Proportion Correct		
	S2 Sample [a]	S3 Sample [b]
Non SD/LEP	0.51 [c]	
SD/LEP ASSESSED	0.31	0.30
EXCLUDED	0	

[a] Sample tested with accommodations not permitted
[b] Sample tested with accommodations permitted
[c] Because the non-SD/LEP students were used in both reporting samples, the original and recomputed means for these students include students from both the S2 and S3 samples.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Delaware Sampling Conclusions

There were no problems found with the sampling of Delaware students. This was investigated in two ways. First, an expert sampling statistician reviewed the 2002 sampling for Delaware and concluded that there were no problems; inclusion of all Delaware schools led to increased accuracy and did not in and of itself increase or decrease score estimates. Second, the weighted count of students from the NAEP sample was closely comparable to enrollment counts from the Delaware Department of Education.

On the other hand, close inspection of Delaware data by subsample, race/ethnicity, and SD/LEP category revealed some apparent inconsistency. One might speculate that there was a shift in exclusion policy between the S2 and S3 sample but that is only speculation. On the other hand, there were small sample sizes in the Hispanic categories. Recognizing the small sample sizes for Hispanics, Table 2.16 repeats the means and gains for Delaware Hispanics for Grades 4 and 8, this time with standard error and confidence interval data. The lower and upper confidence bounds represent a 95-percent confidence level, that is, one can state with 95-percent certainty that the true gain fell within this range. The confidence interval for the Grade 4 Reading gain, for example, ranges from 13 to 59 — a very wide range. Note that the Grade 4 confidence interval does not extend down to zero; therefore, one can conclude with confidence that there was, indeed, an increase in performance between the two years. This is consistent with NCES analyses that show a statistically significant gain for Delaware Hispanics.[2] Thus, the data suggest that Delaware Grade 4 Hispanics did gain, but they also indicate that the confidence interval is very wide due to small sample sizes and large standard error.

Table 2.16. Standard Errors for Delaware Hispanic Reading Gains

	1998 (Recomputed)		2002		1998-2002 Gain		Gain
	Mean	SE	Mean	SE	Gain	SE	95% Confidence Interval
Grade 4	176	11.6	212	1.9	36	11.8	+13 – +59
Grade 8	248	7.9	250	2.1	2	8.2	-14 – +18

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Standard errors for the means and gains, plus the 95-percent confidence interval for the gains, are reported for Delaware as a whole in Table 2.17. As expected, because of the larger sample sizes standard errors of the gains for all Delaware students are smaller (less than 2) and the confidence intervals are narrower than for the Hispanic analyses presented in Table 2.16. Table 2.17 reveals that the 95-percent confidence bands are above 0 for both Grade 4 and Grade 8, indicating that Delaware student performance did improve in 2002, relative to 1998. Indeed, for each grade the lower bound of the confidence interval is above 10.

Table 2.17. Standard Errors for Delaware Reading Gains: All Students

	1998 (Recomputed)		2002		1998-2002 Gain		Gain
	Mean	SE	Mean	SE	Gain	SE	95% Confidence Interval
Grade 4	207	1.7	224	.61	17	1.8	+13 – +21
Grade 8	254	1.3	268	.69	14	1.5	+11 – +17

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

[2] Personal communication (Taslima Rahman, NCES, April 22, 2003)

CHAPTER 3: WEIGHTING

Question 2: Was there a problem with the weighting [case weights] of the Delaware data?

We detected no problem with the case weights of the Delaware data. Delaware is one of the few states where every school is sampled, and nearly all of the students in the targeted grades were tested in 2002. Consequently, the sampling weight assigned to each school should be 1.0. This was, in fact, the case for the 2002 data file.

Further, the student weights should all be the same except for minor reassignment of the weights assigned to students who were absent and not tested. The review of the 2002 student weights found them to be entirely consistent with what is known about the sampling design and results.

Finally, as reported in Chapter 2, we compared the weighted counts from the NAEP reading data file to Delaware Department of Education population statistics. Given transfers in and out of the state after NAEP testing and a few students who may have been retained in grade, the NAEP estimate is close to the counts reported by Delaware. Consequently, one can conclude that NAEP did account for all of Delaware's Grade 4 students in the 2002 Reading assessment.

Delaware Weighting Conclusions

No problems were found with the weighting of Delaware data. As expected, the census testing of all schools and all students within those schools in 2002 yielded appropriate weights.

CHAPTER 4: BIB SPIRAL

Question 3: Was there a problem with the design for assigning test booklets to students (BIB spiral)?

No problem with the BIB (balanced incomplete block) spiral was detected. Grade 4 testing used 32 books and Grade 8 testing used 37 books. The distribution of Grade 4 books and items was even across the state. About 3 percent of participating students received each of the 32 books. Each item was administered to approximately 950 to 1,000 students. Books were evenly distributed within schools to the extent possible (i.e., with 32 books it was not possible to have exactly equal use of books within schools other than size 32, 64, etc.).

Delaware Grade 8 test book distribution also is consistent with distribution across the nation as a whole. Across all states, 18.25 percent of Grade 8 students received one particular test book (out of 37 test books).[3] In Delaware, 18.23 percent of 8th graders received that version. Items, on the other hand, were distributed about equally, as expected. Because the predominant use of one book occurs across the nation, this, by itself, should not explain the large Delaware gain.

Tables 4.1 and 4.2 present the distribution of books within Delaware for Grades 4 and 8, respectively. For each test book, the percentage of students within the state who were assigned that book is indicated. In addition, a school-by-school analysis yielded the minimum percentage of students within a school assigned to a given book. A minimum percentage of zero is expected in schools testing fewer than 32 students because at least one book could not be assigned. The maximum percentage of students within a school assigned to each book is also provided. Had any of these percentages been high, that finding would have been a red flag for potential BIB spiraling problems; no unusual percentages were found.

Delaware BIB Spiraling Conclusions

No problems with the distribution of test books were found, either within any school or across schools in the state. The distribution of books in Delaware schools closely matched the distribution in other states.

[3] Steve Lazer (ETS) indicated that the book distribution (with one book used for almost 20 percent of all students) was intentional. (Personal communication [Steve Lazer, ETS, April 2, 2003])

Table 4.1. Distribution of 2002 4th Grade Reading Books in Delaware

Book ID	Percentage of Delaware Students per Book	Minimum Percentage Within School	Maximum Percentage Within School
1	3.2	0.0	11.1
2	3.1	0.0	9.1
3	3.3	0.0	10.0
4	2.9	0.0	11.1
5	3.0	0.0	11.1
6	3.0	0.0	11.1
7	3.2	0.0	9.1
8	3.2	0.0	7.1
9	3.2	0.0	11.1
10	3.2	0.0	9.1
11	3.1	0.0	7.1
12	3.1	0.0	7.1
13	3.1	0.0	7.1
14	2.9	0.0	7.1
15	3.3	0.0	7.1
16	3.2	0.0	7.1
17	3.1	0.0	7.1
18	3.2	0.0	6.8
19	3.2	0.0	6.8
20	3.1	0.0	10.0
21	3.3	0.0	10.0
22	3.3	0.0	10.0
23	3.2	0.0	10.0
24	3.1	0.0	10.0
25	3.4	0.0	10.0
26	3.1	0.0	6.3
27	3.0	0.0	10.0
28	3.0	0.0	11.1
29	3.3	0.0	11.1
30	3.3	0.0	11.1
31	3.0	0.0	11.1
32	2.8	0.0	5.4

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 4.2. Distribution of 2002 8th Grade Reading Books in Delaware

Book ID	Percentage of Delaware Students per Book	Minimum Percentage Within School	Maximum Percentage Within School
1	2.2	0.0	3.7
2	2.5	0.0	3.9
3	2.4	0.0	3.7
4	2.4	0.0	3.9
5	2.4	0.0	3.9
6	2.0	0.0	3.9
7	2.2	0.0	3.9
8	2.2	0.0	4.2
9	2.3	0.0	4.2
10	2.4	0.0	4.2
11	2.3	0.0	7.4
12	2.4	0.0	4.2
13	2.3	0.0	4.2
14	2.2	0.0	14.3
15	2.3	0.0	14.3
16	2.2	0.0	14.3
17	2.2	0.0	14.3
18	2.3	0.0	14.3
19	2.2	0.0	14.3
20	2.1	0.0	4.2
21	2.2	0.0	4.8
22	2.1	0.0	4.2
23	2.4	0.0	4.8
24	2.2	0.0	4.8
25	2.2	0.0	4.8
26	2.3	0.0	4.8
27	2.4	0.0	4.8
28	2.3	0.0	4.8
29	2.2	0.0	3.9
30	2.4	0.0	4.8
31	2.3	0.0	4.8
32	2.4	0.0	4.8
33	2.4	0.0	4.8
34	2.3	0.0	4.8
35	2.1	0.0	4.8
36	2.4	0.0	3.9
37	18.2	14.3	22.1

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

CHAPTER 5: SCORING

Question 4: Was there a problem with the scoring (hand scoring or scanning) of the Delaware data?

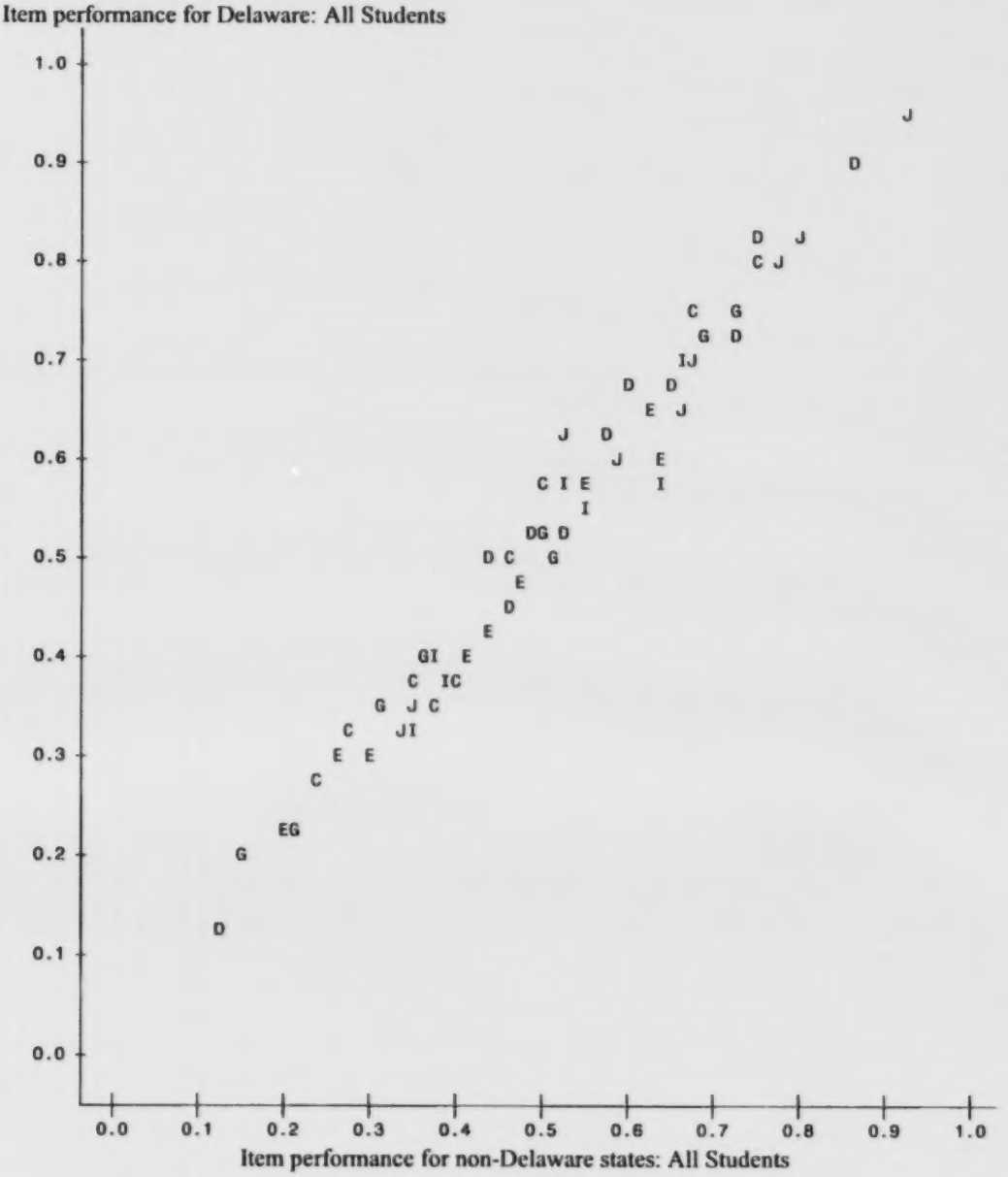
According to Pearson, all responses to any single open-ended item are scored at the same time in the same scoring location with student responses from different states randomly distributed during the scoring process. Therefore, systematic bias in scoring that would cause Delaware scores to be too high seems unlikely. However, HumRRO produced a variety of scatterplot diagrams relating item performance for Delaware in comparison with the rest of the nation. These plots should show a reasonably tight, elliptically-slanted pattern indicative of a strong correlation between item performance (and its inverse, item difficulty) in Delaware and the rest of the nation. Items or sets of items that do not fall in the diagonal pattern may signal unexpected scoring problems, coding problems, breaches of security, or exposure problems.

Figures 5.1 through 5.4, on the following pages, show the relationship between item performance in Delaware and the rest of the nation for Grade 4 Reading for all students and by race/ethnicity. Multiple-choice item performance is simply the p-value, or the proportion of students who answered the item correctly. To put open-ended performance on an analogous 0-to-1 scale, item mean performance was divided by total possible points. Items are labeled by block; the blocks correspond to different reading passages.

Figure 5.1 shows tightly clustered items, as expected, and reveals no apparent pattern to the arrangement of the letters by passage in these plots. The pattern is less tightly clustered for Hispanic students (Figure 5.2), as would be expected because of the smaller sample size. The cluster appears more tightly packed for Blacks (Figure 5.3) and even tighter for Whites (Figure 5.4), which, again, is consistent with larger sample sizes.

Figures 5.5 through 5.8 are analogous plots for Grade 8. Again, Figure 5.5, showing all students, is tightly clustered, with no apparent pattern to blocks of items. None of the separate plots of each racial/ethnic category (Figures 5.6, 5.7, and 5.8) reveals anything unexpected.

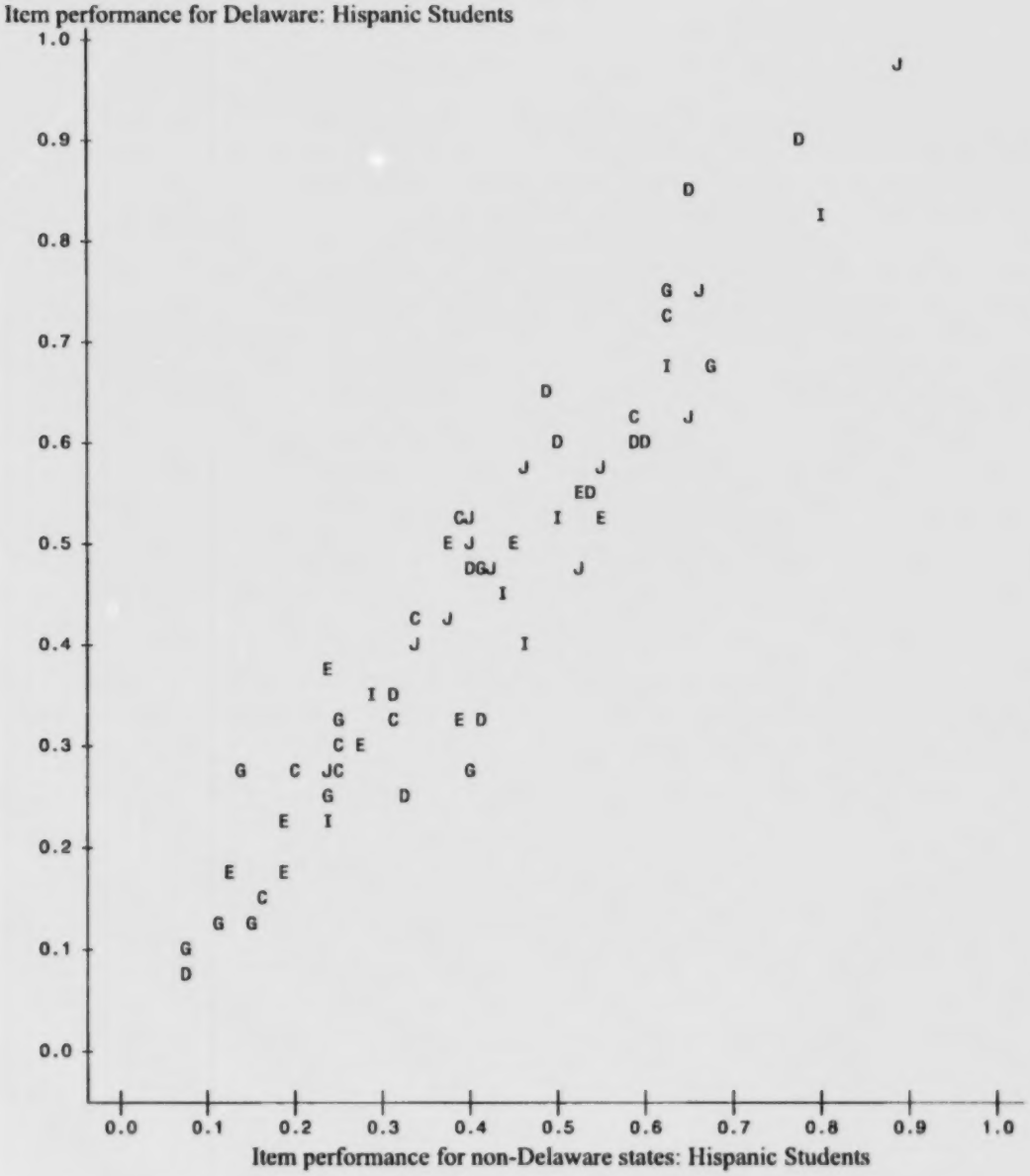
Figure 5.1. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for all students



NOTE: Letter symbol identifies block. Nine observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

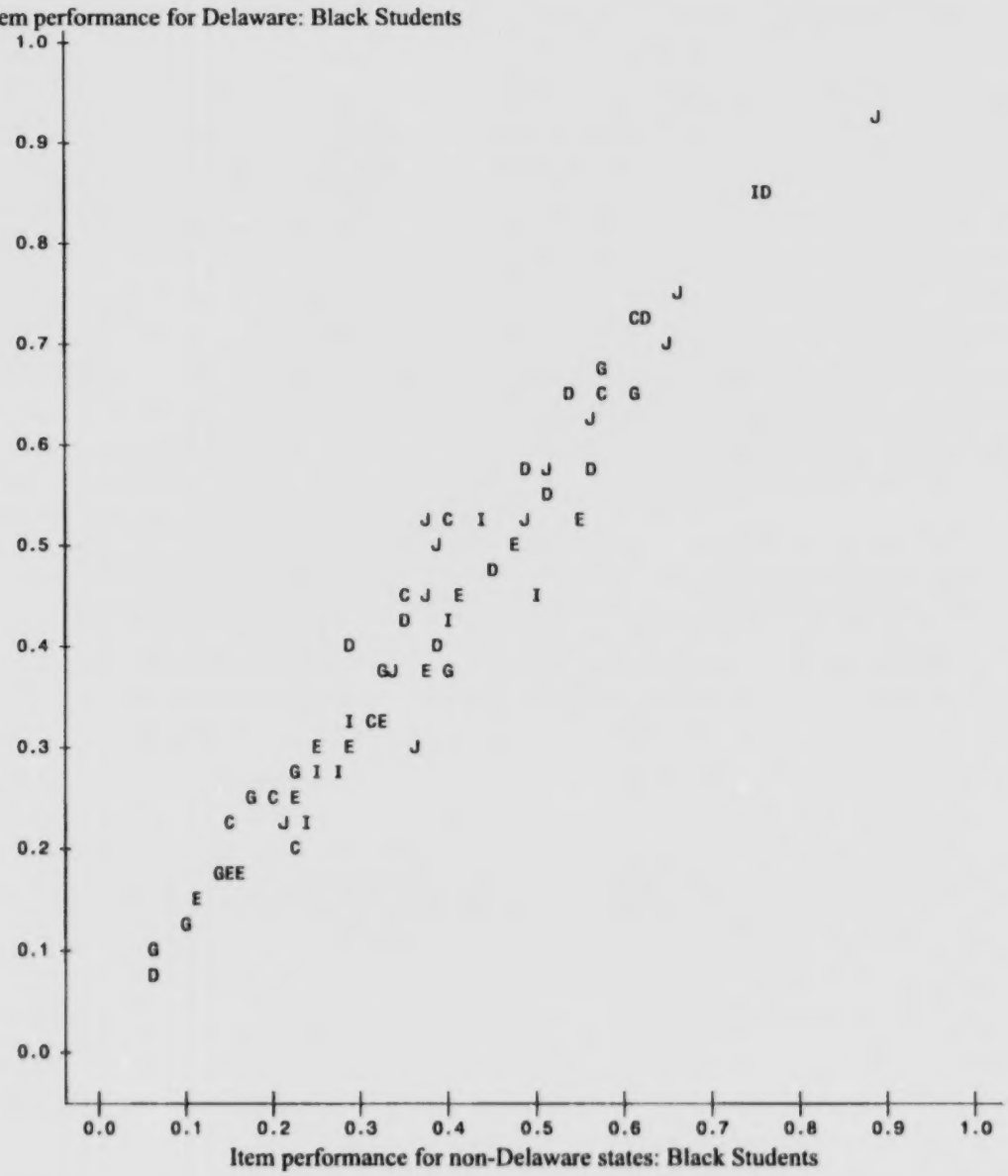
Figure 5.2. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Hispanic students



NOTE: Letter symbol identifies block. Three observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

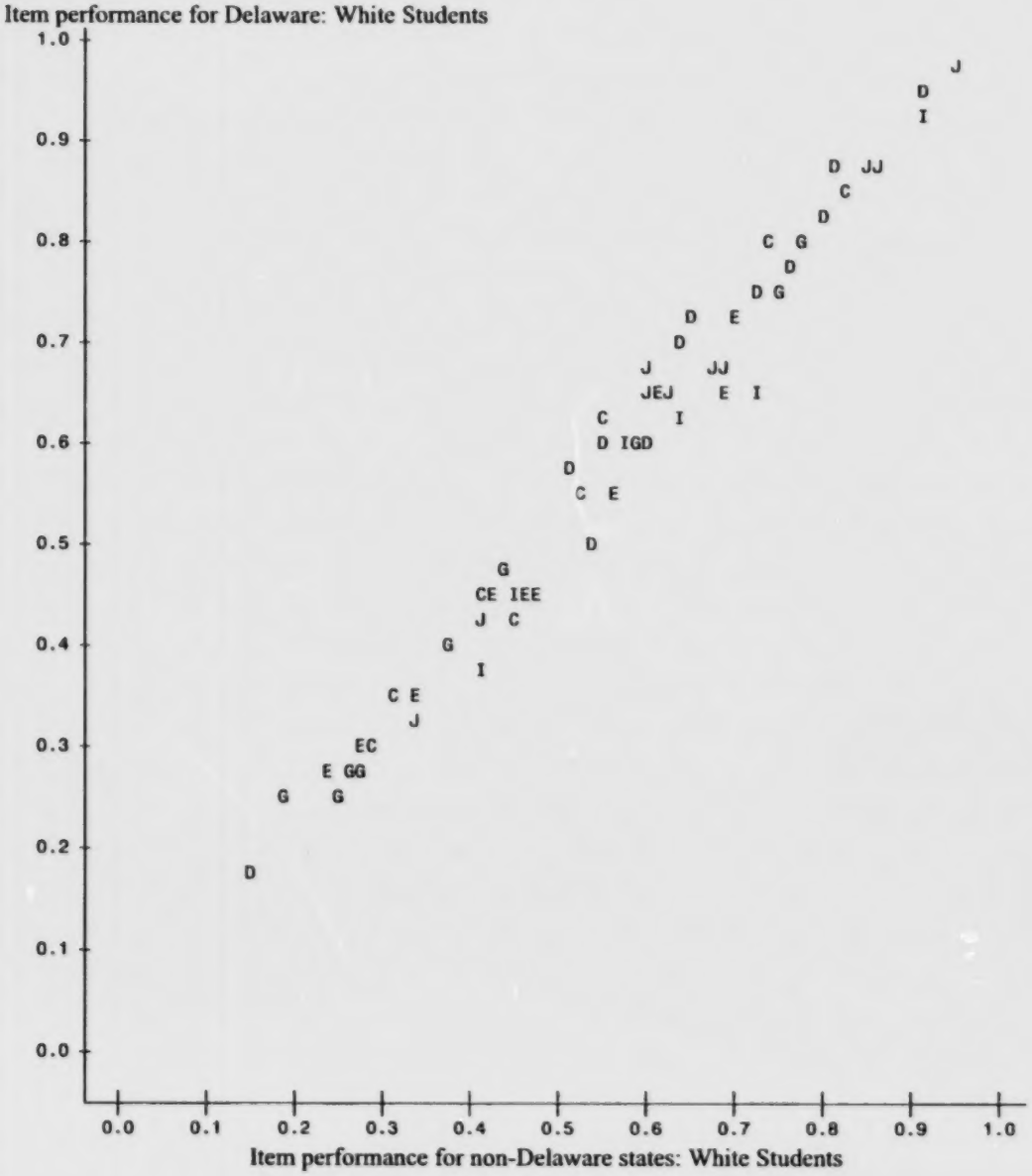
Figure 5.3. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Black students



NOTE: Letter symbol identifies block. Four observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

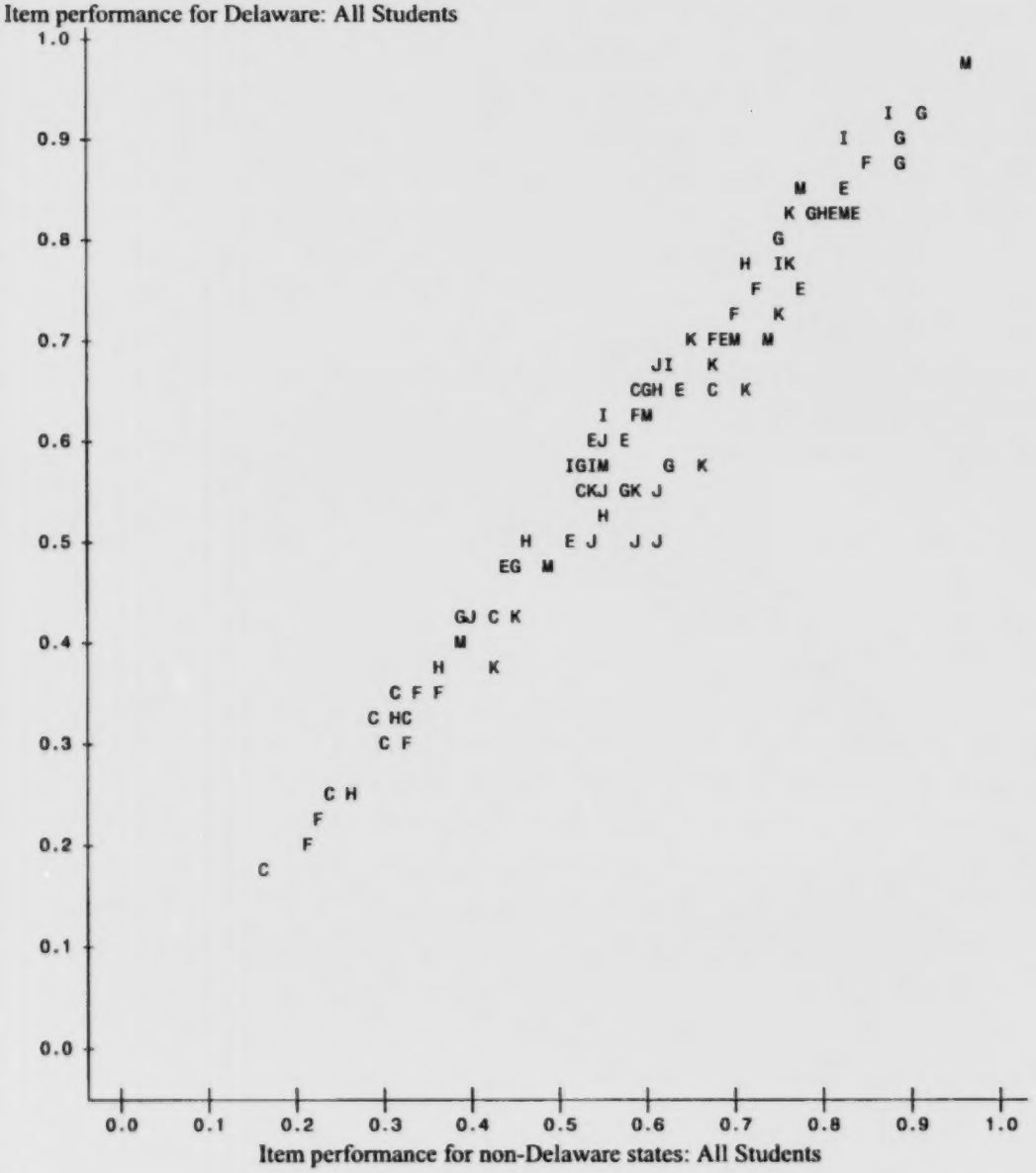
Figure 5.4. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for White students



NOTE: Letter symbol identifies block. Eight observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

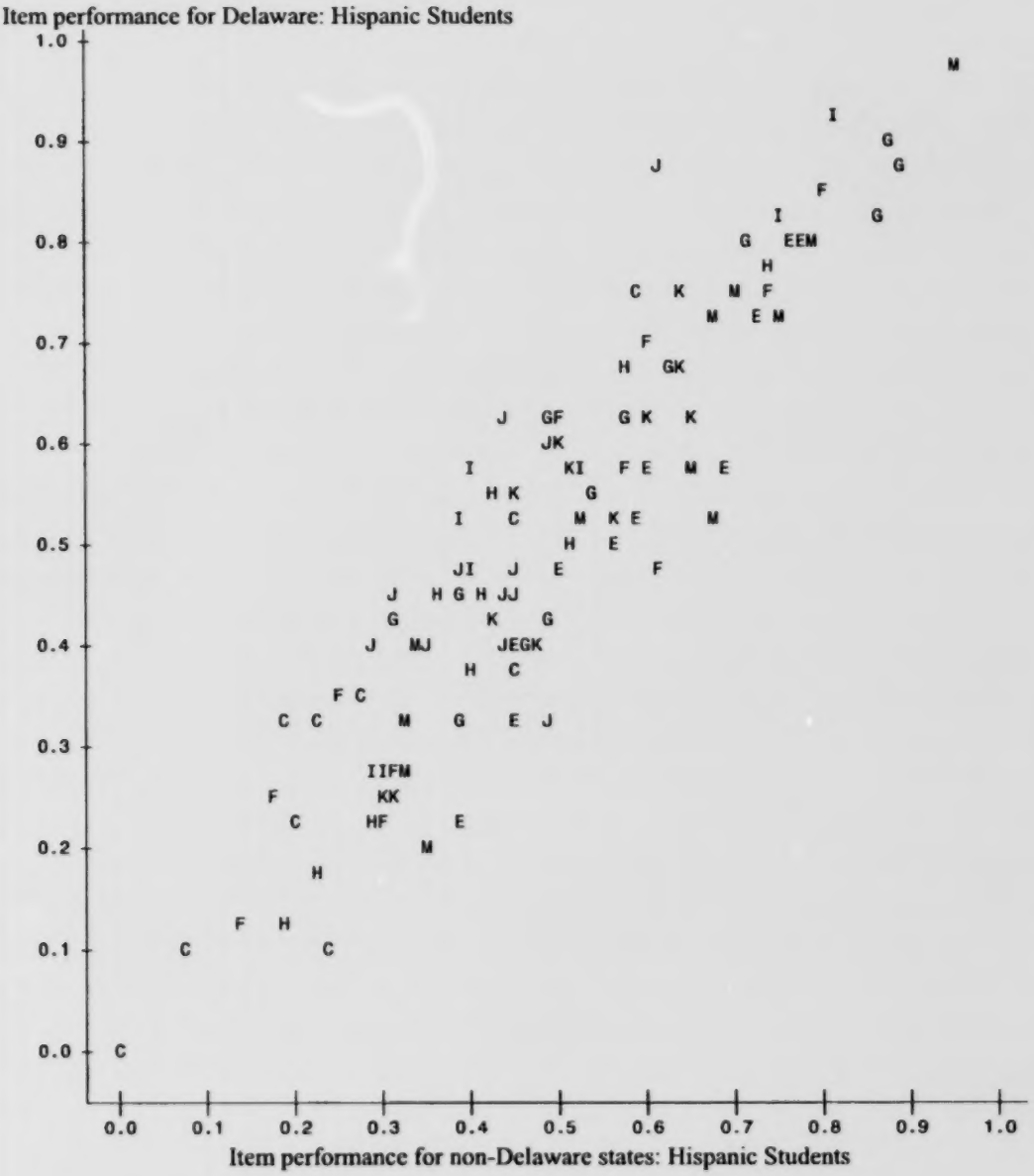
Figure 5.5. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for all students



NOTE: Letter symbol identifies block. Eighteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

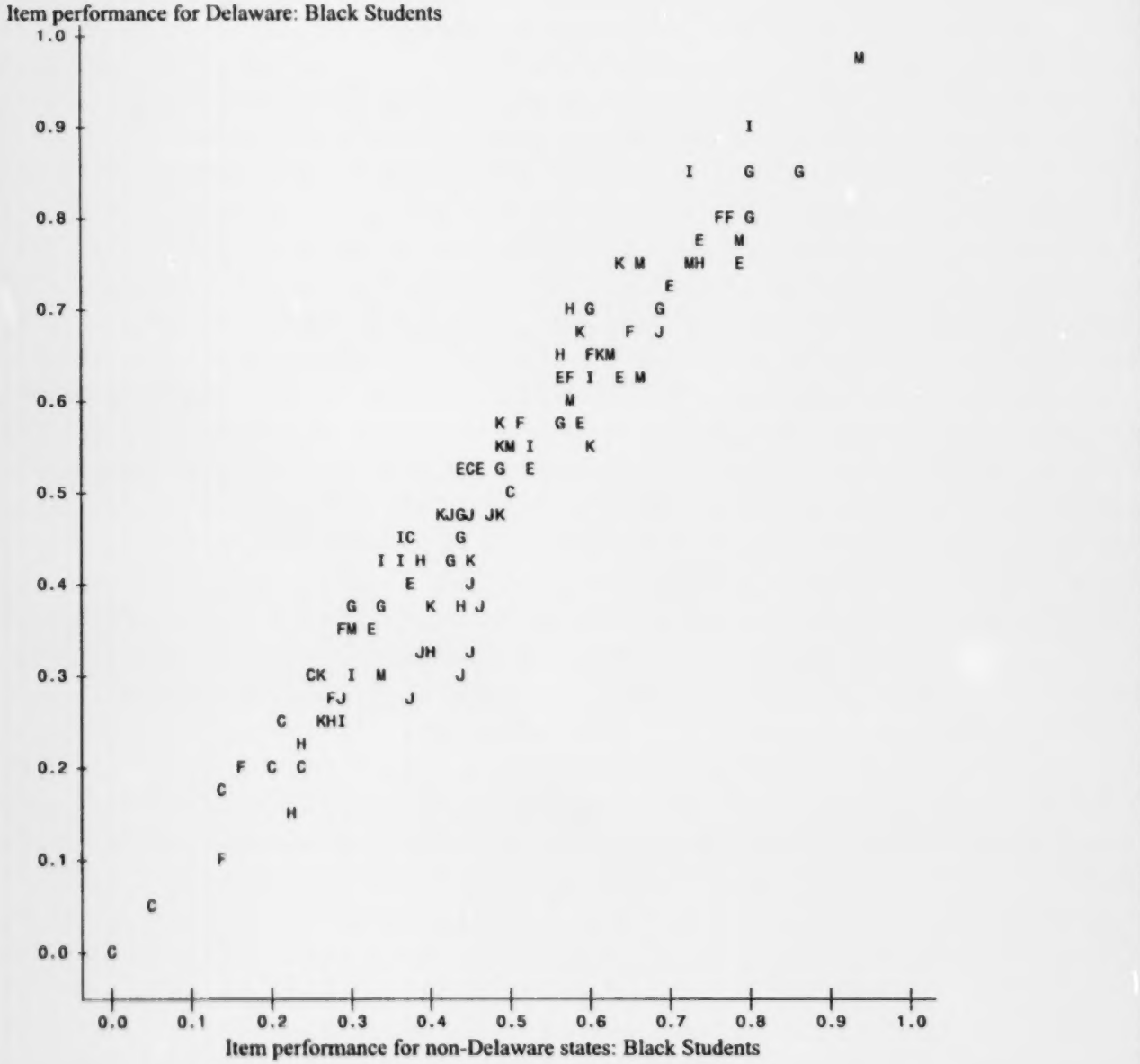
Figure 5.6. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Hispanic students



NOTE: Letter symbol identifies block. Three observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

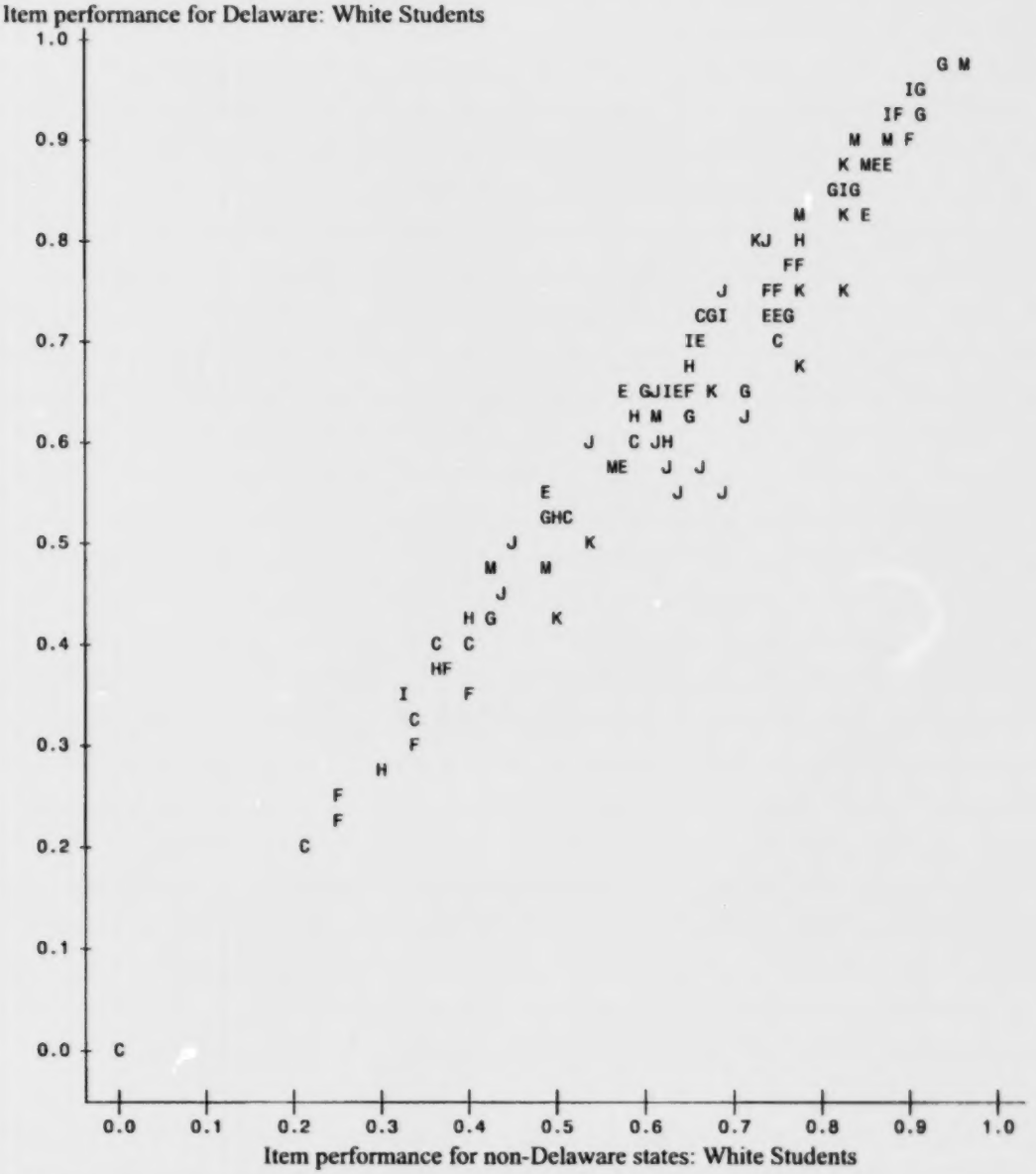
Figure 5.7. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Black students



NOTE: Letter symbol identifies block. Eight observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 5.8. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for White students



NOTE: Letter symbol identifies block. Fourteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

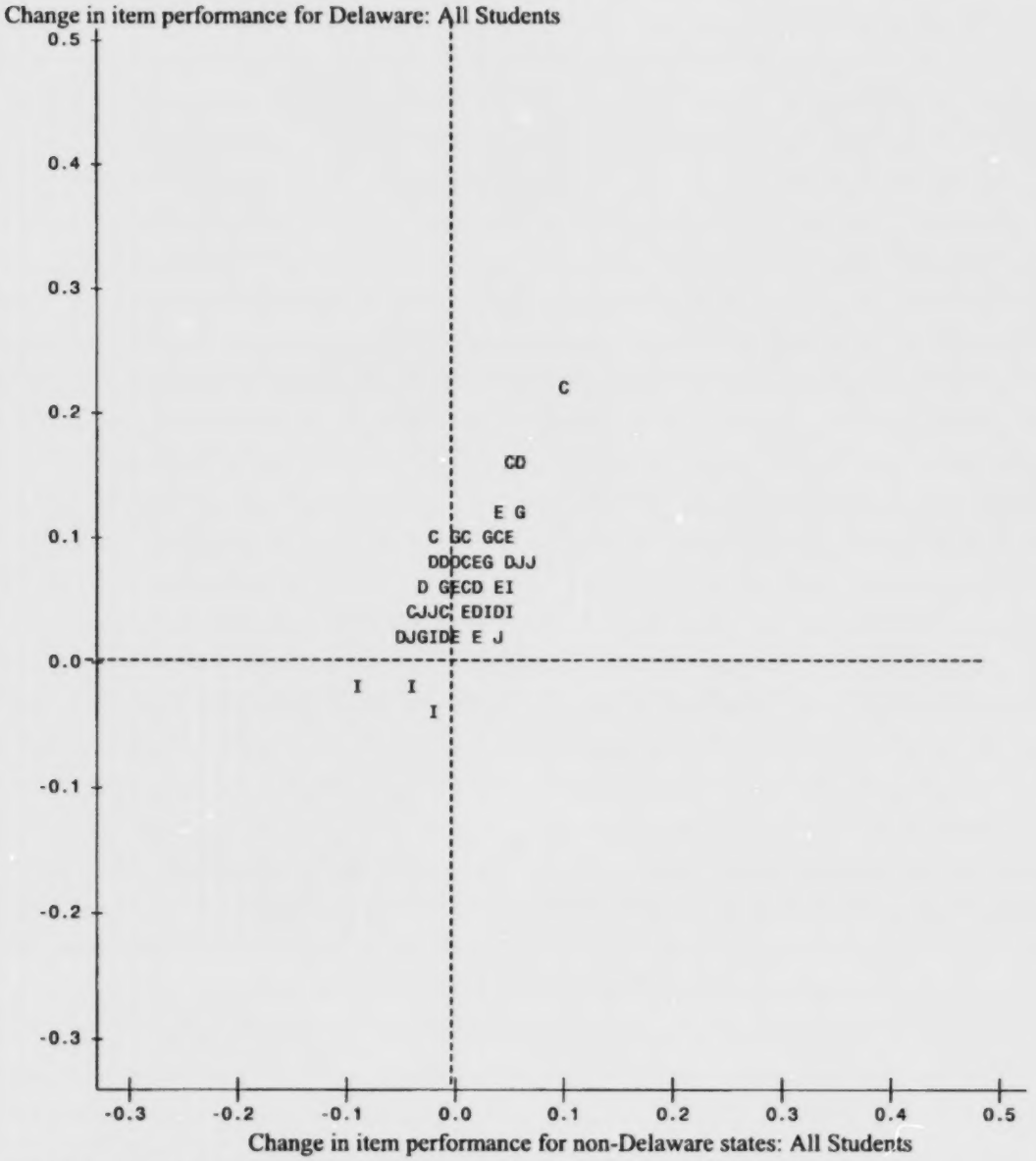
Next, changes in item performance from 1998 to 2002 for Delaware versus non-Delaware states were plotted for a more sensitive view of potential breaches in security. For example, unexpected gains associated with particular item passages could signal teaching to that particular passage. As expected, the plots are more scattered and indicate that Delaware tends to have large gains, but the changes do not appear to be associated with particular blocks of items. Figures 5.9 though 5.12 present the data for all students and each ethnicity category for Grade 4. Figures 5.13 through 5.16 present the data for Grade 8.

HumRRO received information from NCES that Delaware's state test was patterned after NAEP and included both multiple-choice and open-ended items. HumRRO's experience in two other states that also use both types of items prompted speculation that Delaware students may receive special instruction to facilitate their performance on the open-ended items. Therefore, all the plots presented below were repeated, this time with the items labeled by "M" for multiple-choice or "O" for open-ended. This exploration proved to be informative.

Figures 5.17 though 5.20 show Grade 4 relationships between item performance for Delaware compared to the rest of the nation. Looking closely at Figure 5.17, the Os (encircled separately from the Ms) do appear to perform differently. First, the ellipse enclosing the open-ended items is lower and to the left of the ellipse for the multiple-choice items, indicating that the (adjusted) mean performance for the open-ended items is lower than the p-values for the multiple-choice items. This difference may or may not be very important since the lower bound for average performance for multiple-choice items is about .25 because of the potential for answering a multiple-choice item correctly by guessing.

On the other hand, the open-ended items appear to be on the top side on the overall pattern, suggesting that Delaware students did a little better on the open-ended items than on the multiple-choice items, in comparison to non-Delaware students. Looking at the center of each ellipse as a way of portraying average performance also shows that Delaware students were performing higher than non-Delaware students on both open-ended and multiple-choice items. Figures 5.18, 5.19, and 5.20 show that the same is true for Hispanic, Black, and White students.

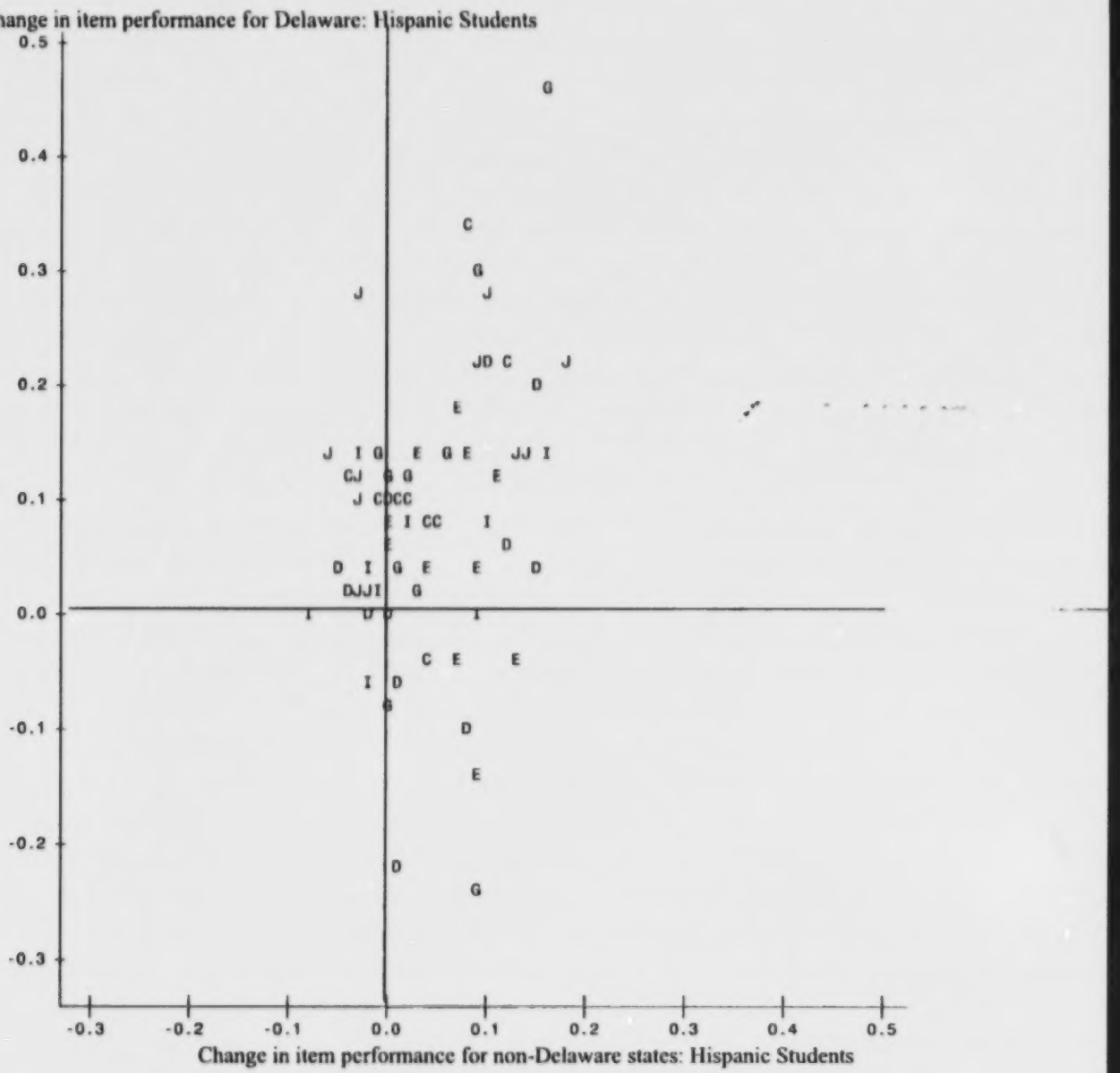
Figure 5.9. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for all students



NOTE: Letter symbol identifies block. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

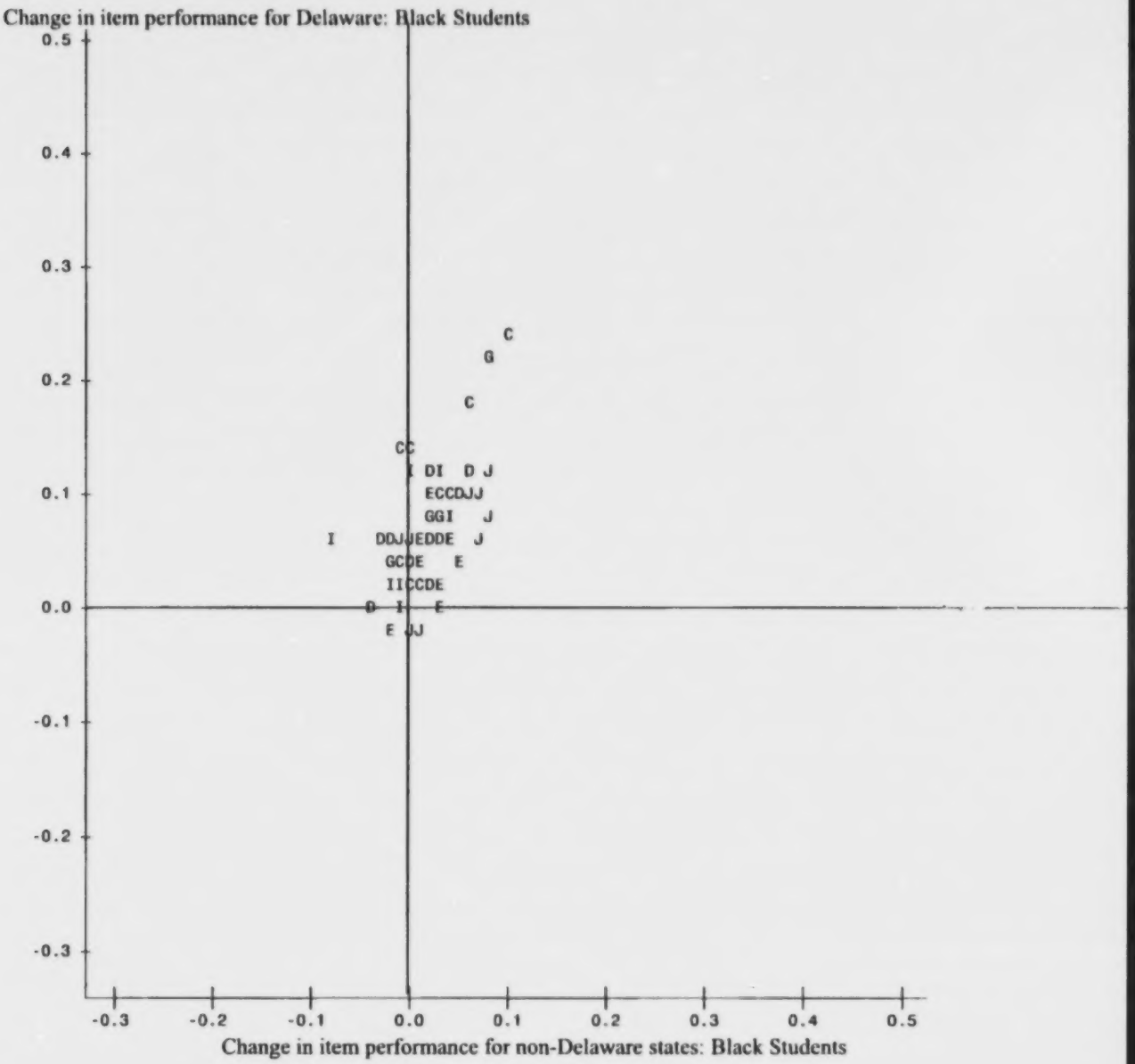
Figure 5.10. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students



NOTE: Letter symbol identifies block. One observation hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

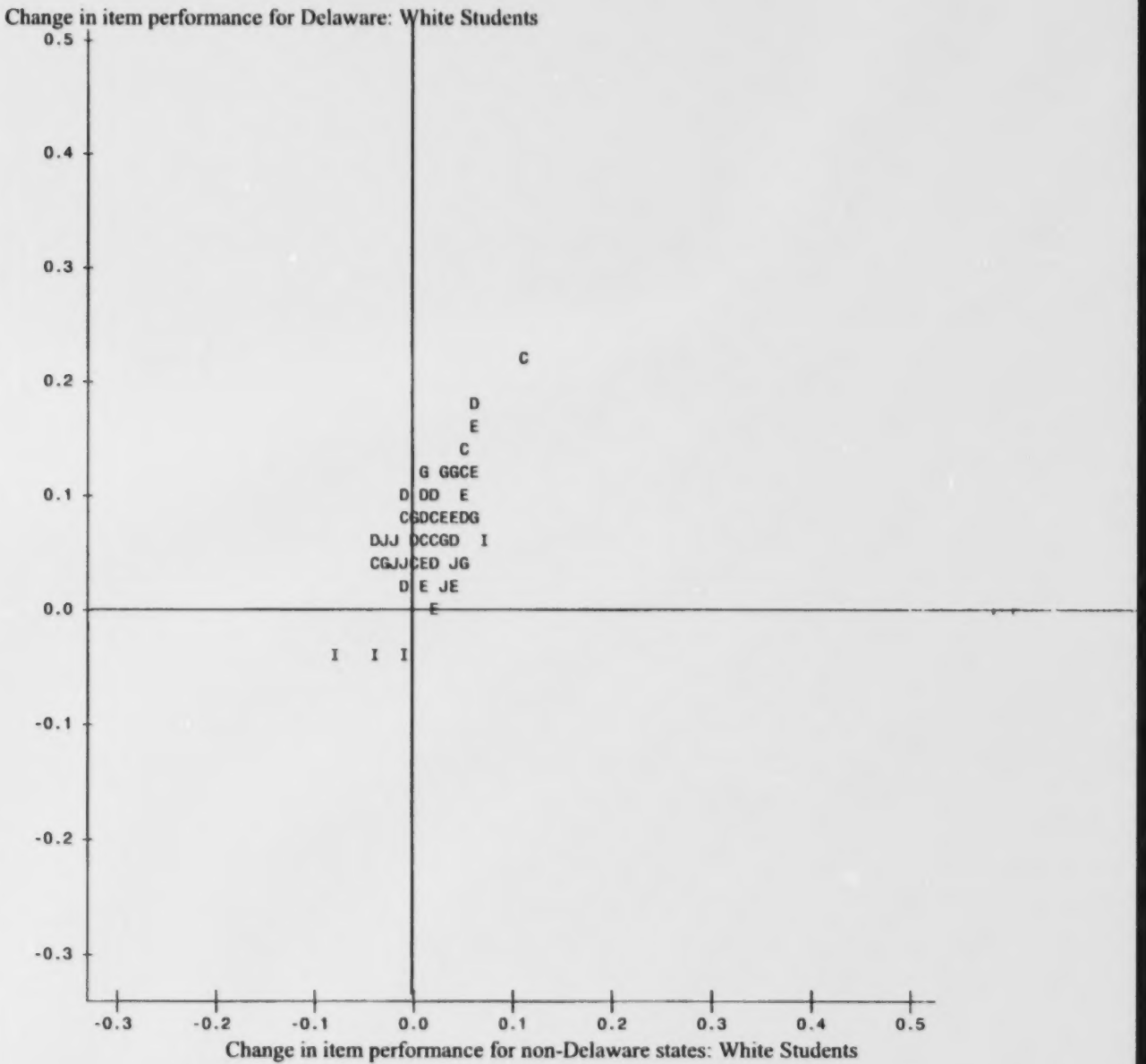
Figure 5.11. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Black students



NOTE: Letter symbol identifies block. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

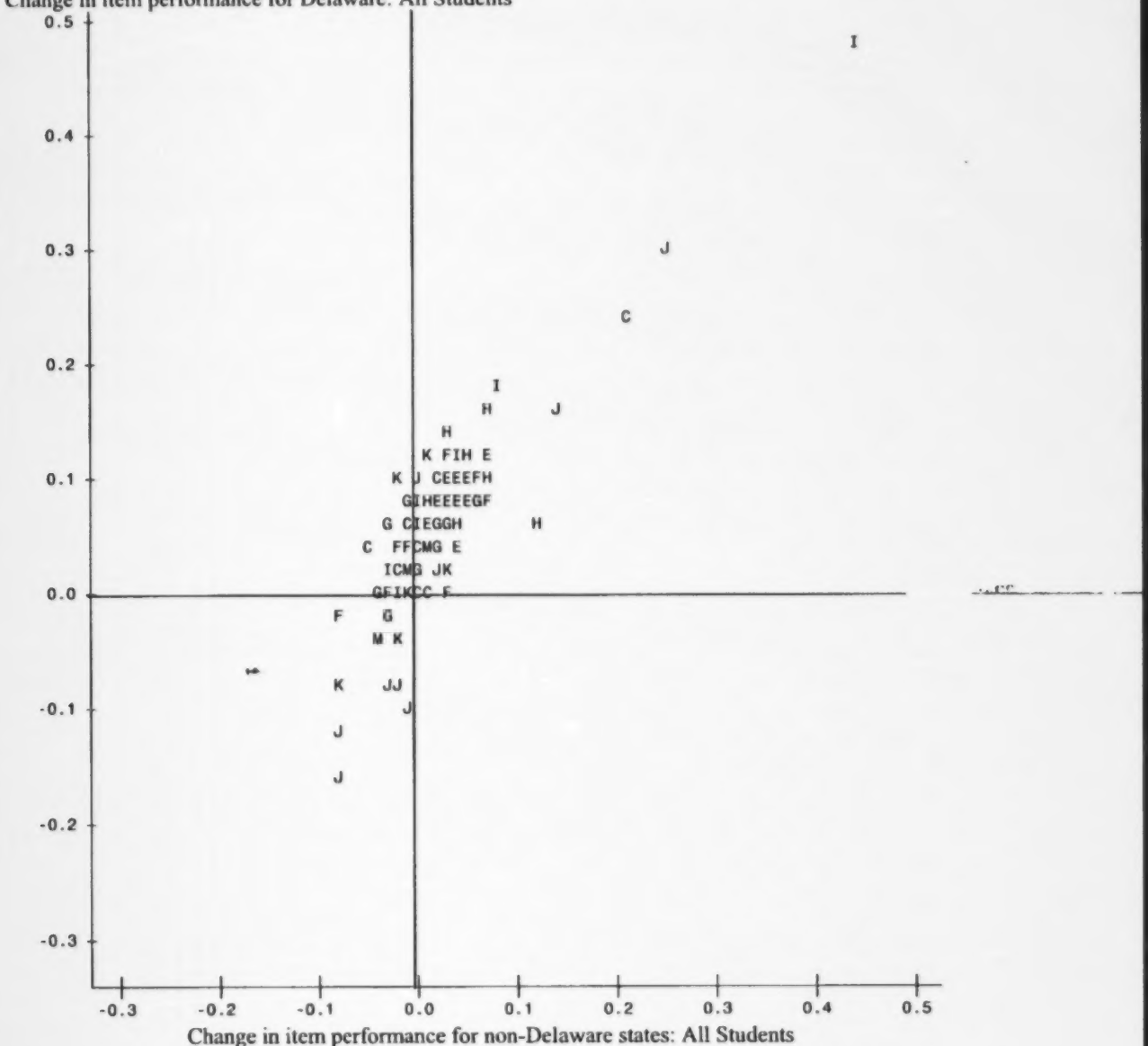
Figure 5.12. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for White students



NOTE: Letter symbol identifies block. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

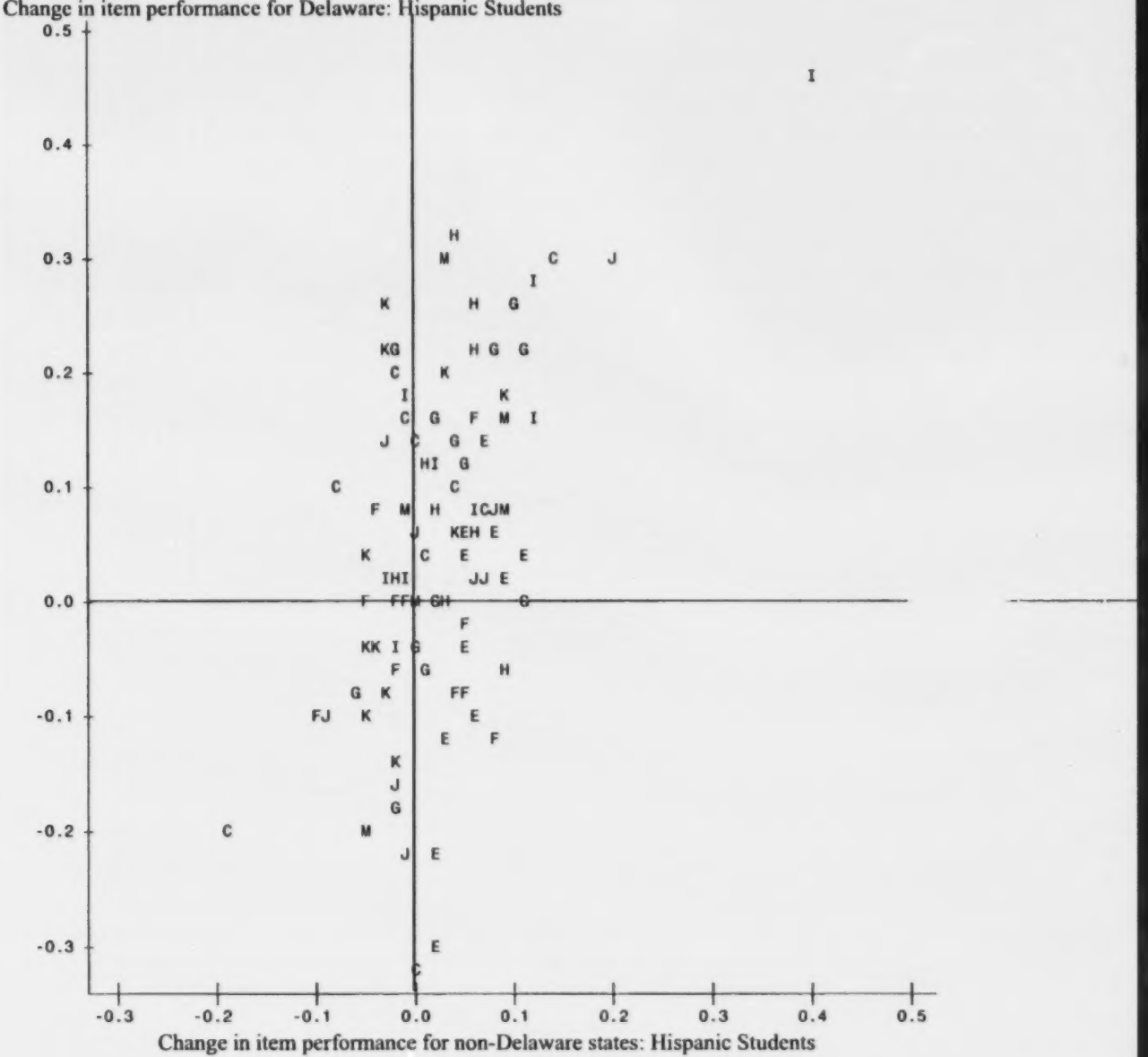
Figure 5.13. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for all students



NOTE: Letter symbol identifies block. Thirty-five observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

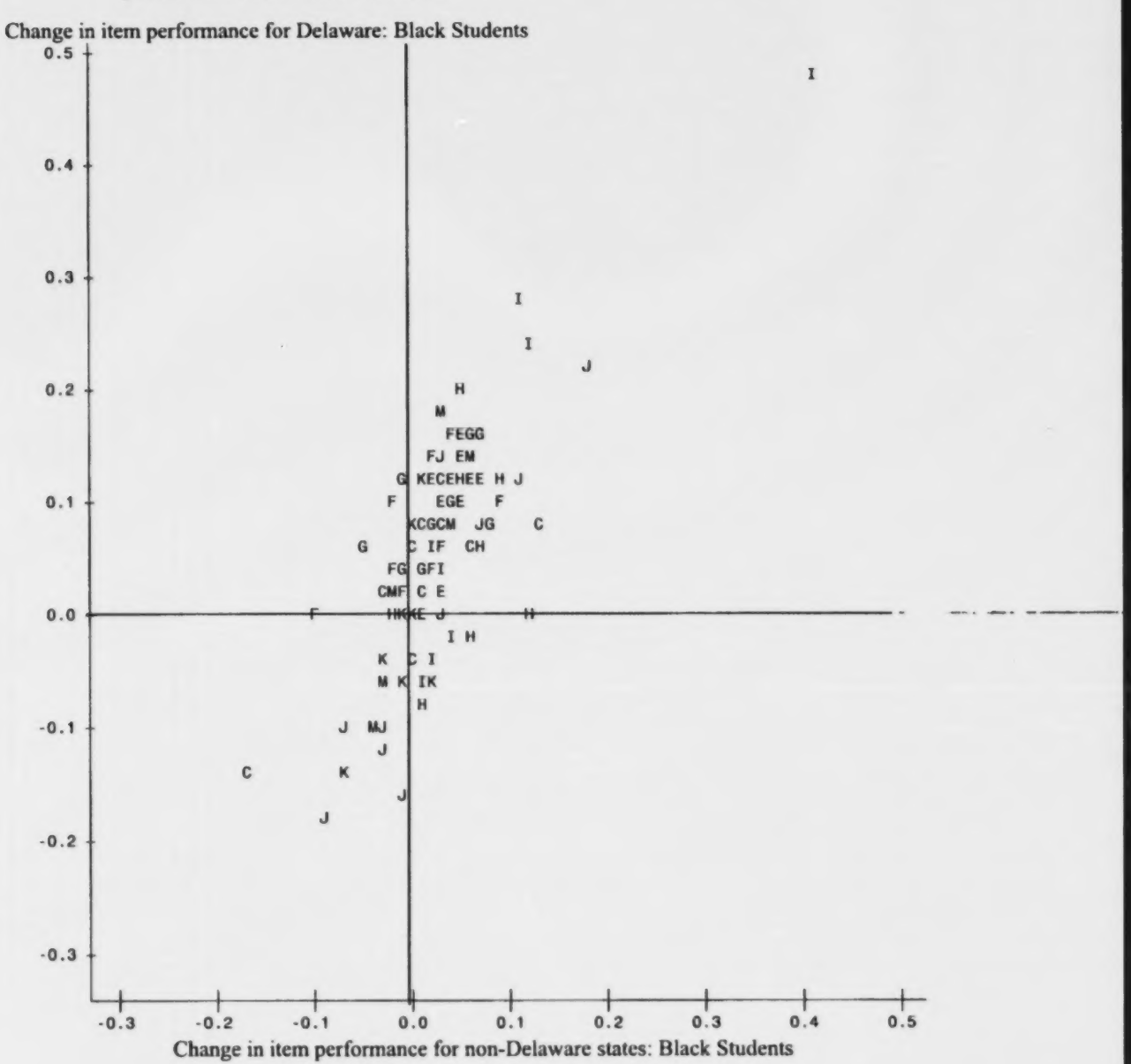
Figure 5.14. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students



NOTE: Letter symbol identifies block. Thirteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

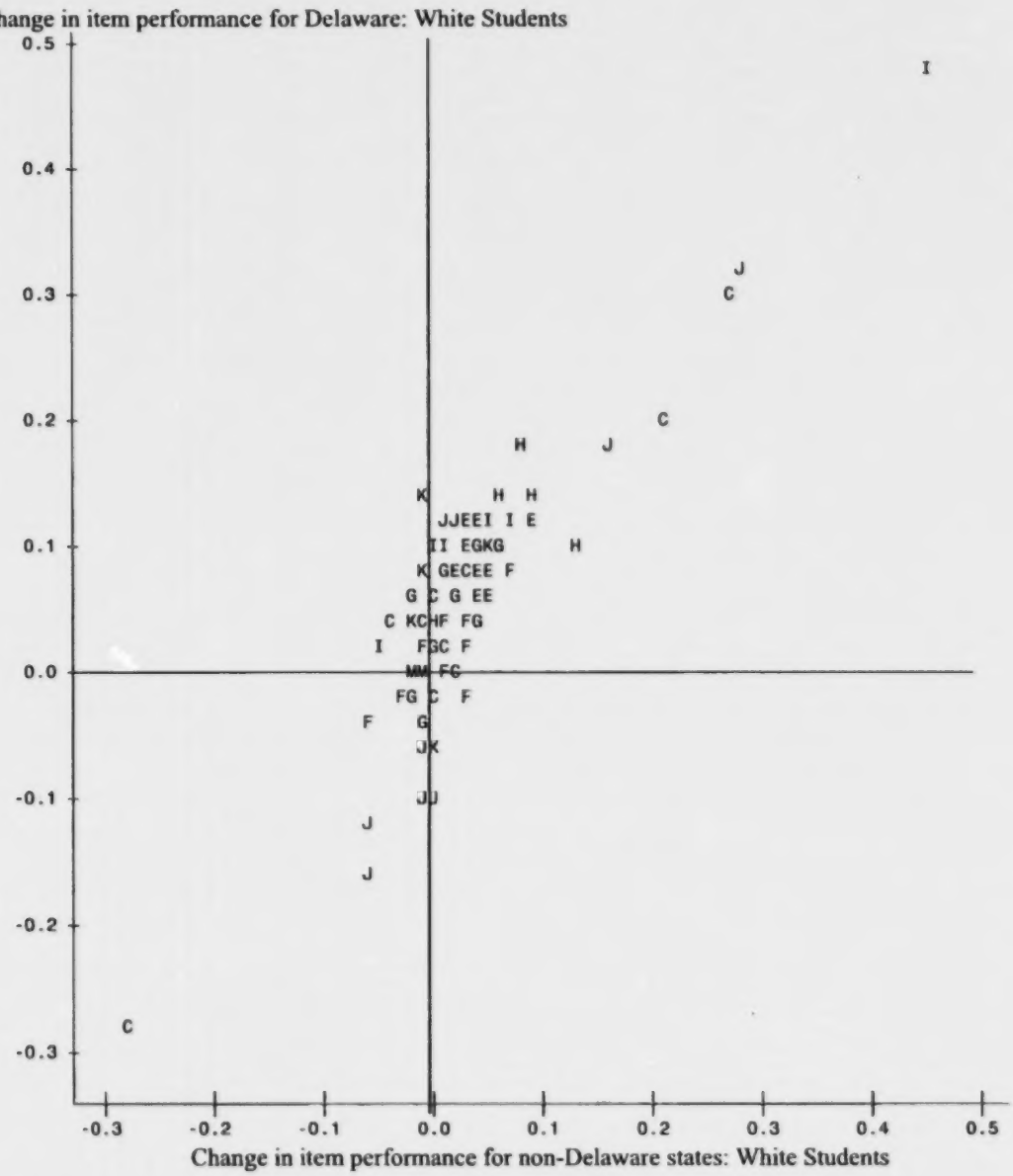
Figure 5.15. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Black students



NOTE: Letter symbol identifies block. Twenty-four observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

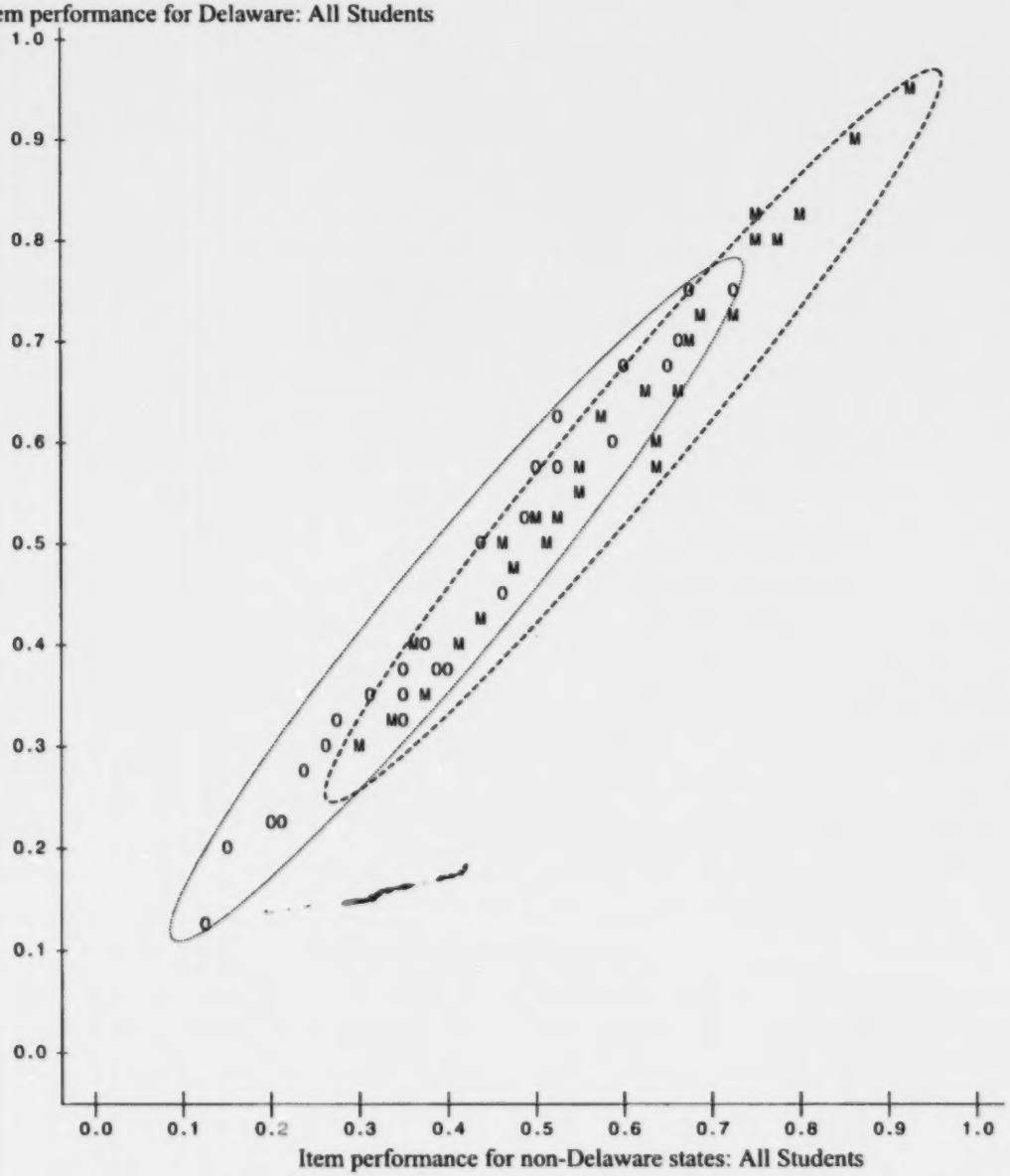
Figure 5.16. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for White students



NOTE: Letter symbol identifies block. Thirty-eight observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

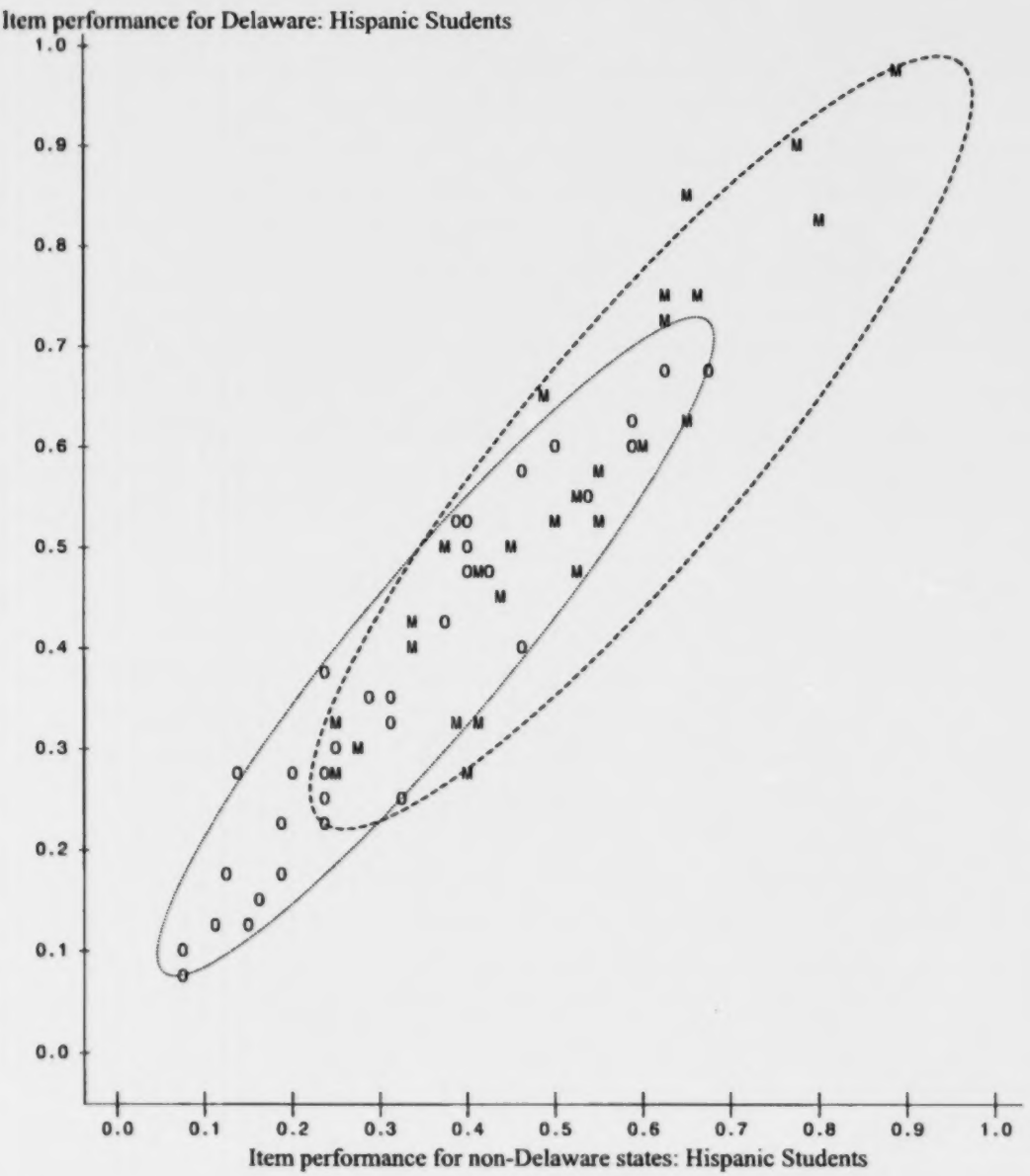
Figure 5.17. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for all students



NOTE: M=multiple choice item; O=open-ended item. Ten observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

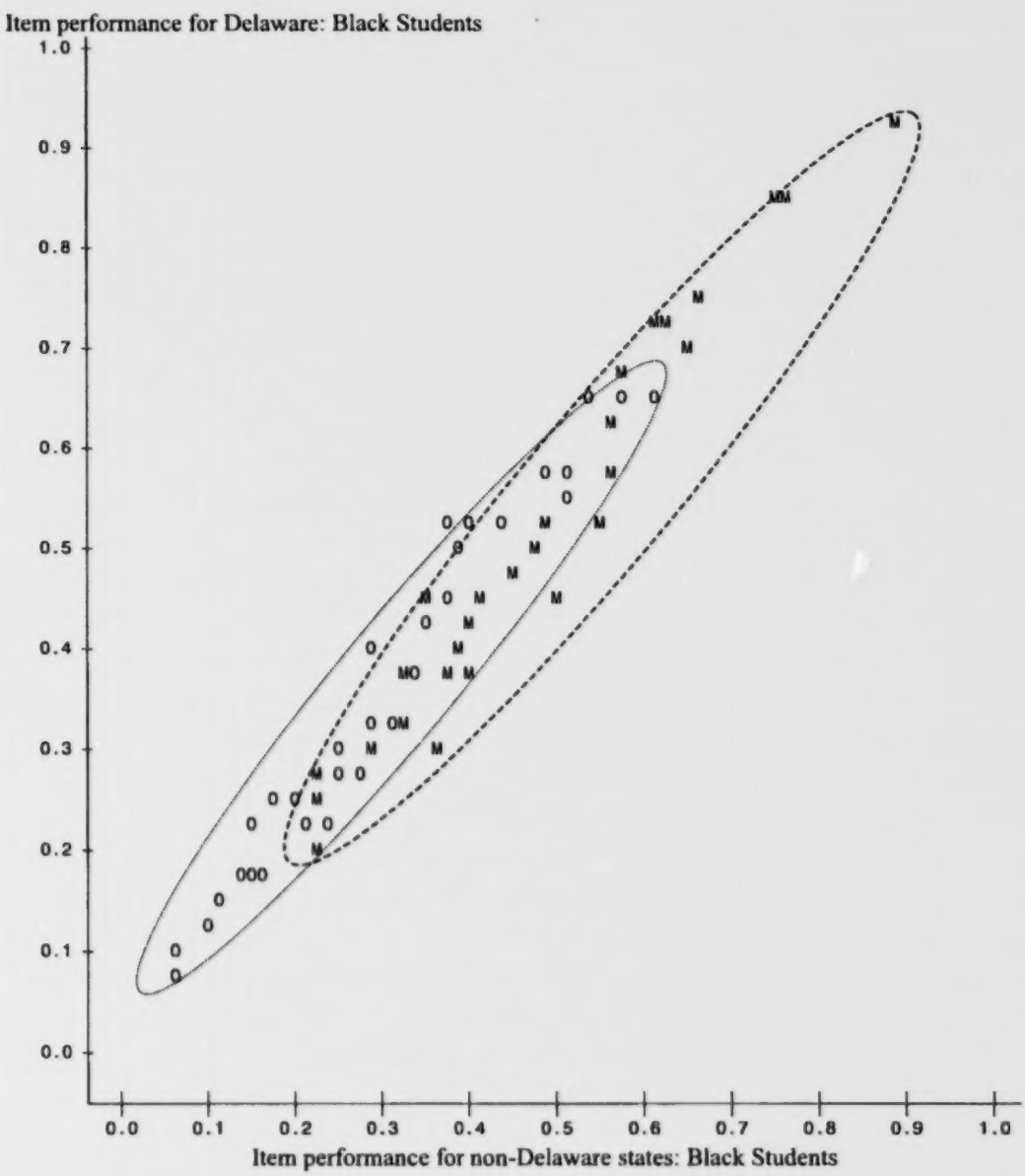
Figure 5.18. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Hispanic students.



NOTE: M=multiple choice item; O=open-ended item. Three observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

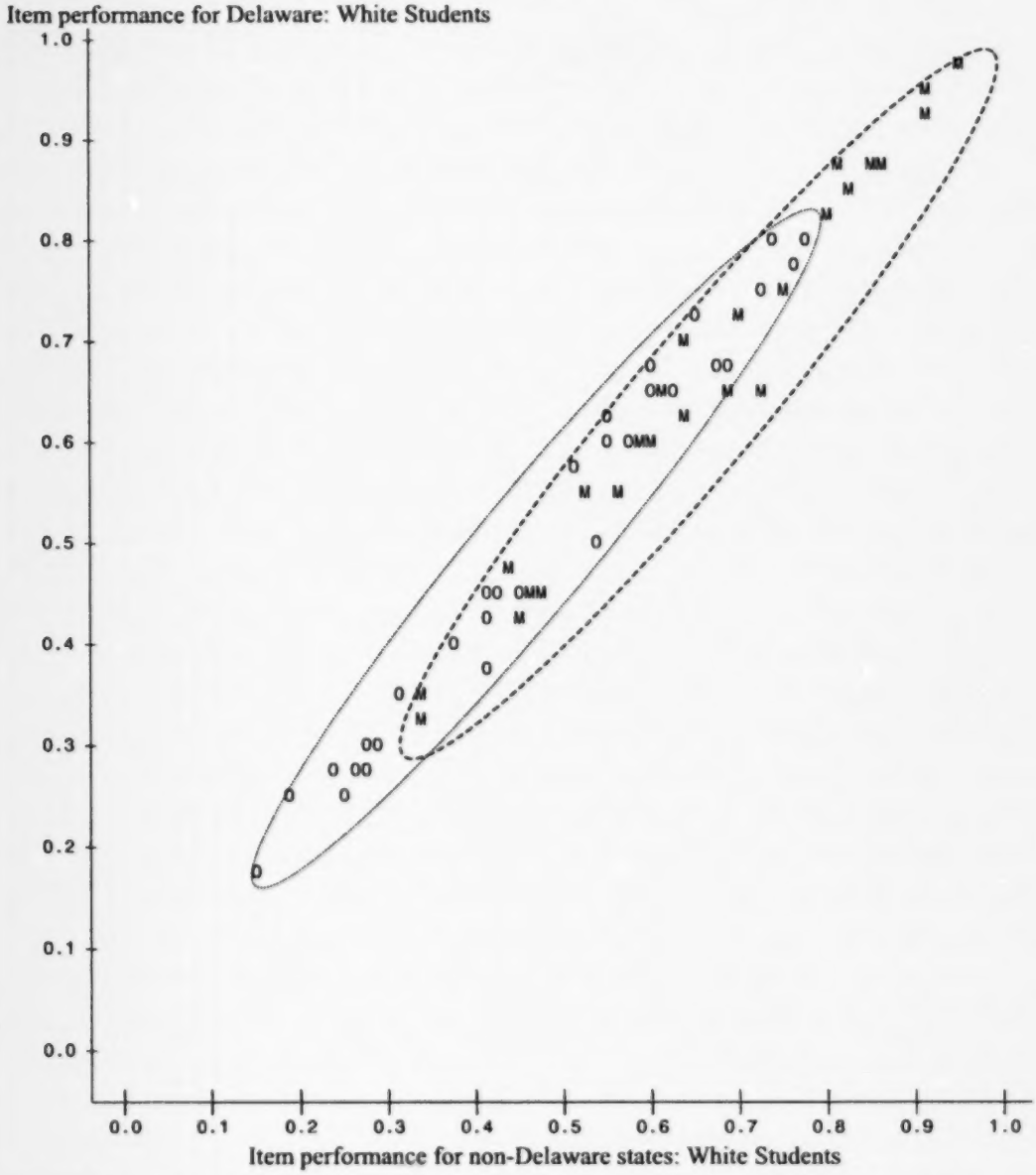
Figure 5.19. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for Black students



NOTE: M=multiple choice item; O=open-ended item. Four observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 5.20. Plot of 2002 Grade 4 non-Delaware states by Delaware item performance for White students



NOTE: M=multiple choice item; O=open-ended item. Eight observations hidden due to overlap.

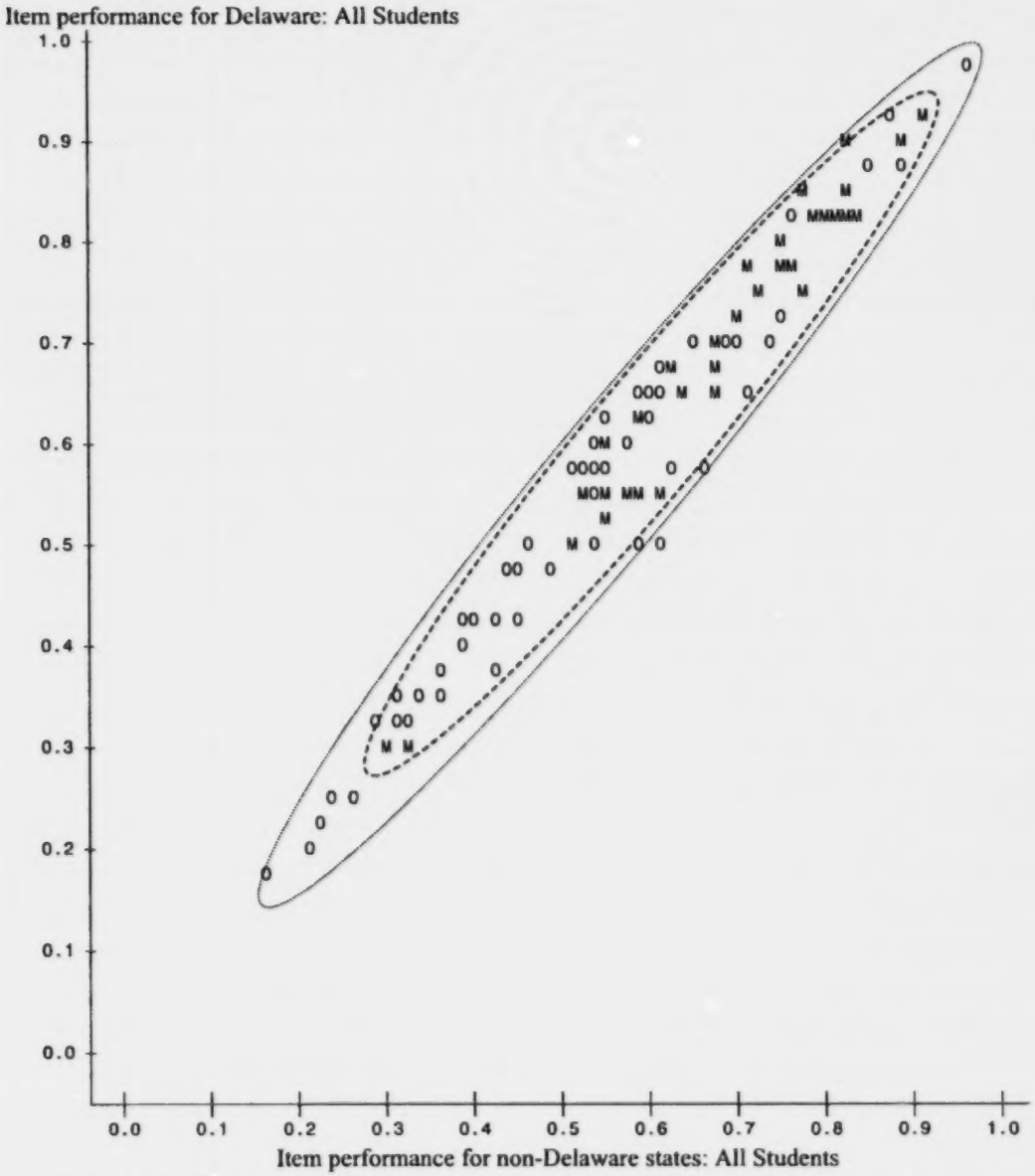
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figures 5.21 through 5.24 show multiple-choice and open-ended performance for Grade 8. The difference between multiple-choice and open-ended items seen in Grade 4 is not apparent in Grade 8.

Finally, we turn to changes in item performance labeled by type of item, presented in Figures 5.25 through 5.28 for Grade 4, and Figures 5.29 through 5.32 for Grade 8. Figure 5.25, the first of these figures, reveals an important finding. It shows that while the nation gained only on multiple-choice items, Delaware improved on both open-ended and multiple-choice items. Therefore, an important part of the difference in score gains between Delaware and the rest of the nation is due to Delaware's gains for open-ended items relative to the rest of the states.

For Grade 8 (see Figure 5.29), the separation of open-ended and multiple-choice item gains is not as dramatic as for Grade 4. However, it does appear that both Grade 8 and Grade 4 improved on multiple-choice items in both Delaware and the rest of the nation, although Delaware appears to have improved a little more. For the open-ended items, Delaware Grade 8 students improved on more of the items then the rest of the nation.

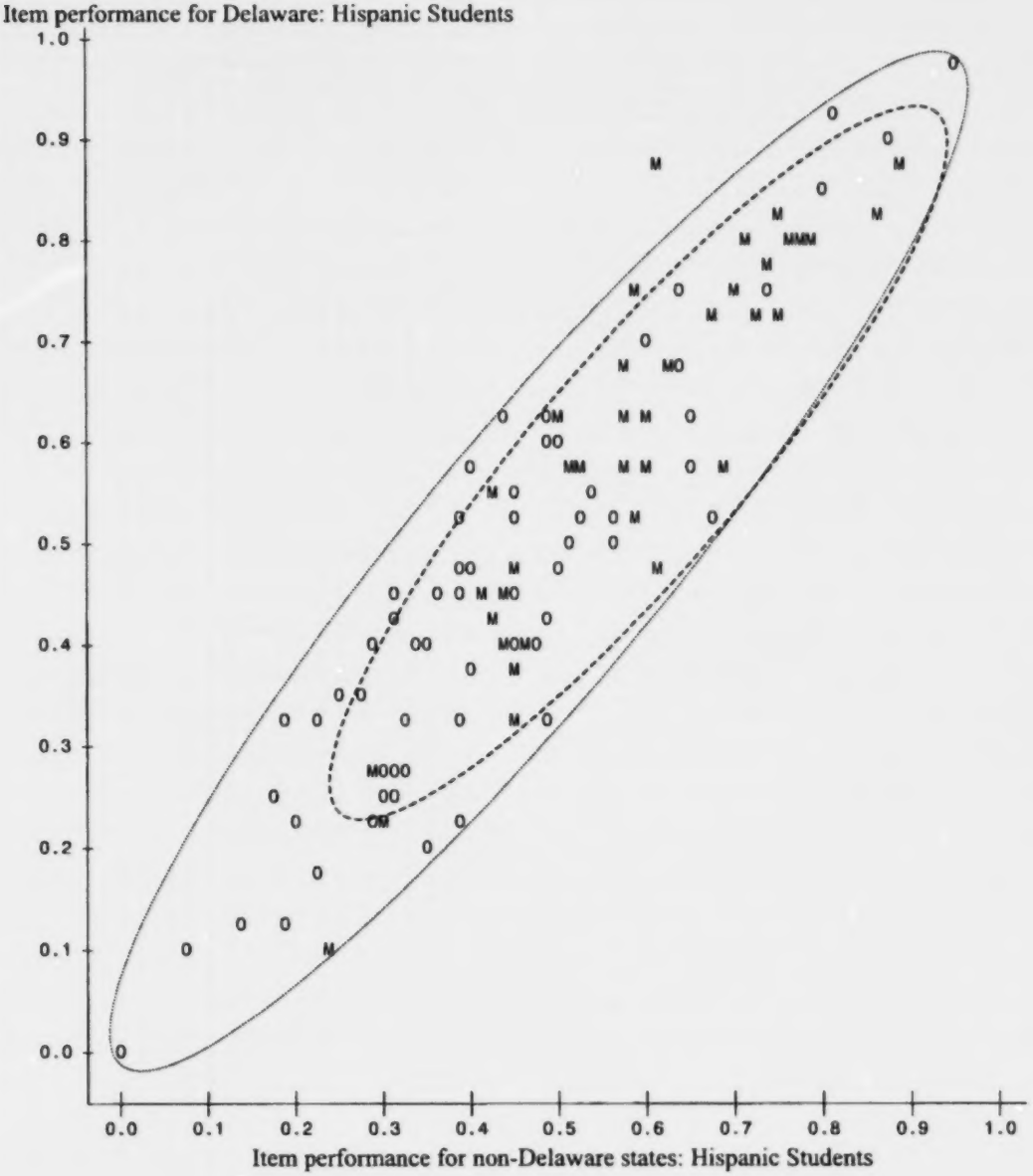
Figure 5.21. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for all students



NOTE: M=multiple choice item; O=open-ended item. Eighteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

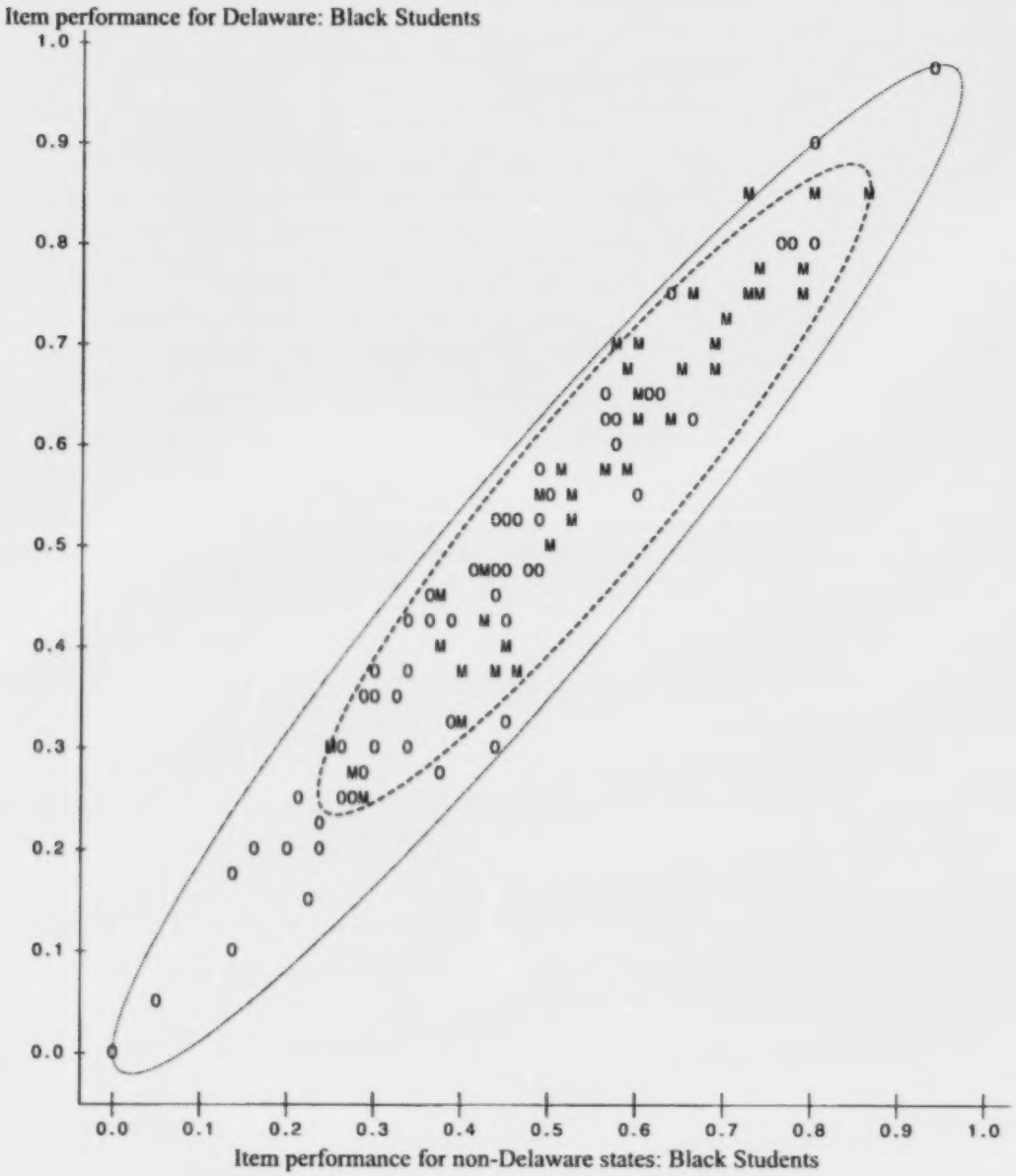
Figure 5.22. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Hispanic students



NOTE: M=multiple choice item; O=open-ended item. Three observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

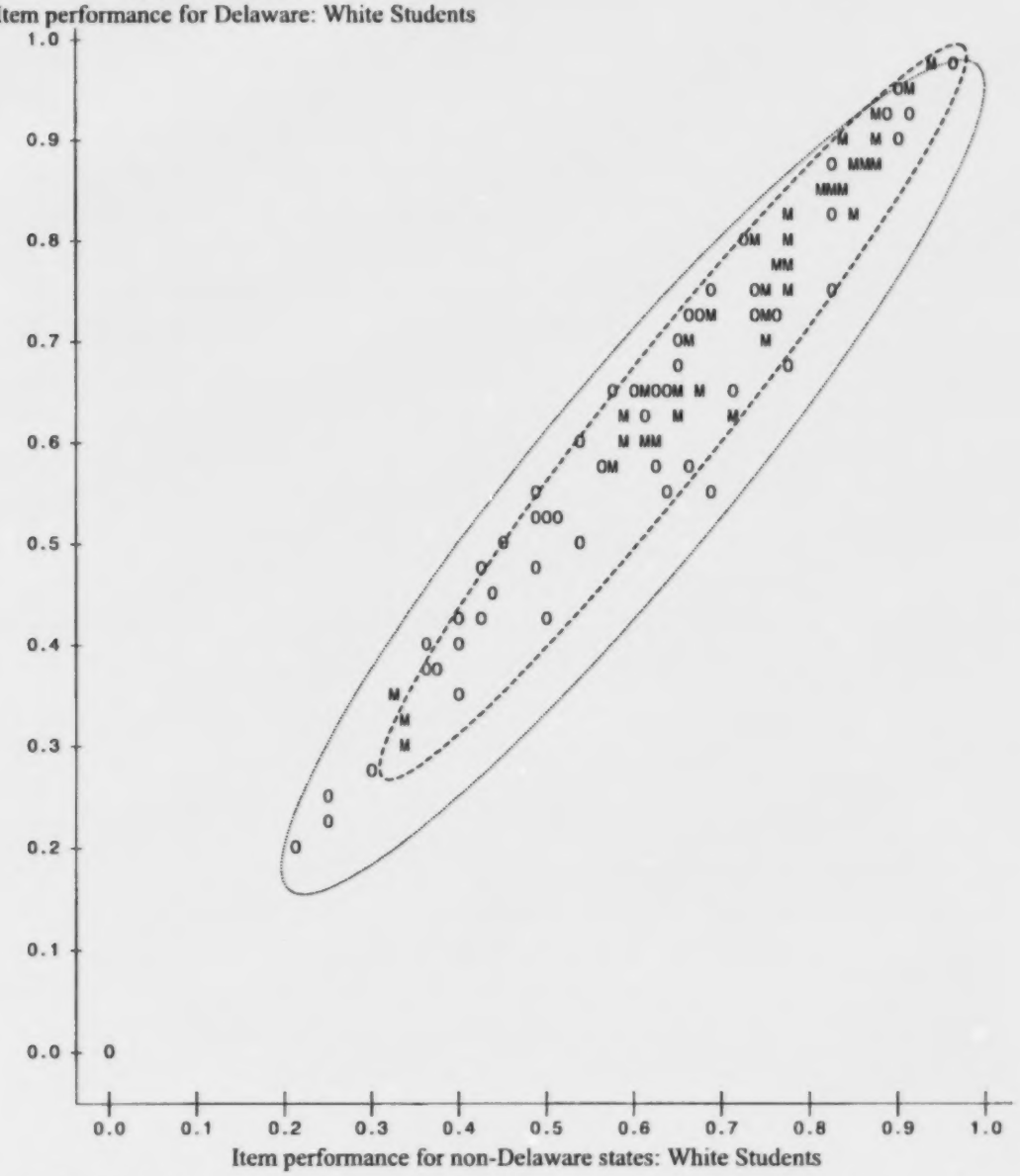
Figure 5.23. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for Black students



NOTE: M=multiple choice item; O=open-ended item. Eight observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

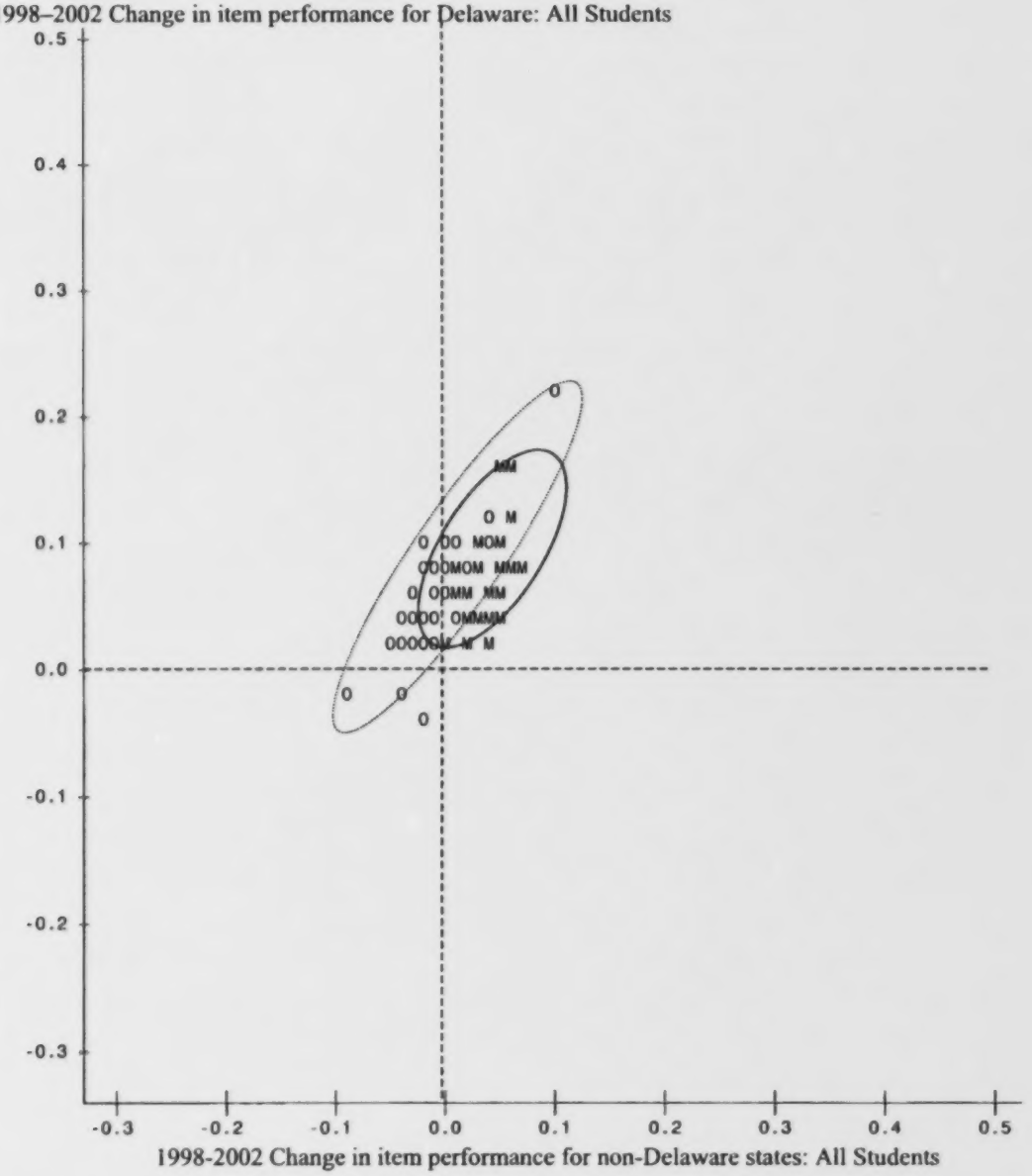
Figure 5.24. Plot of 2002 Grade 8 non-Delaware states by Delaware item performance for White students



NOTE: M=multiple choice item; O=open-ended item. Fourteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

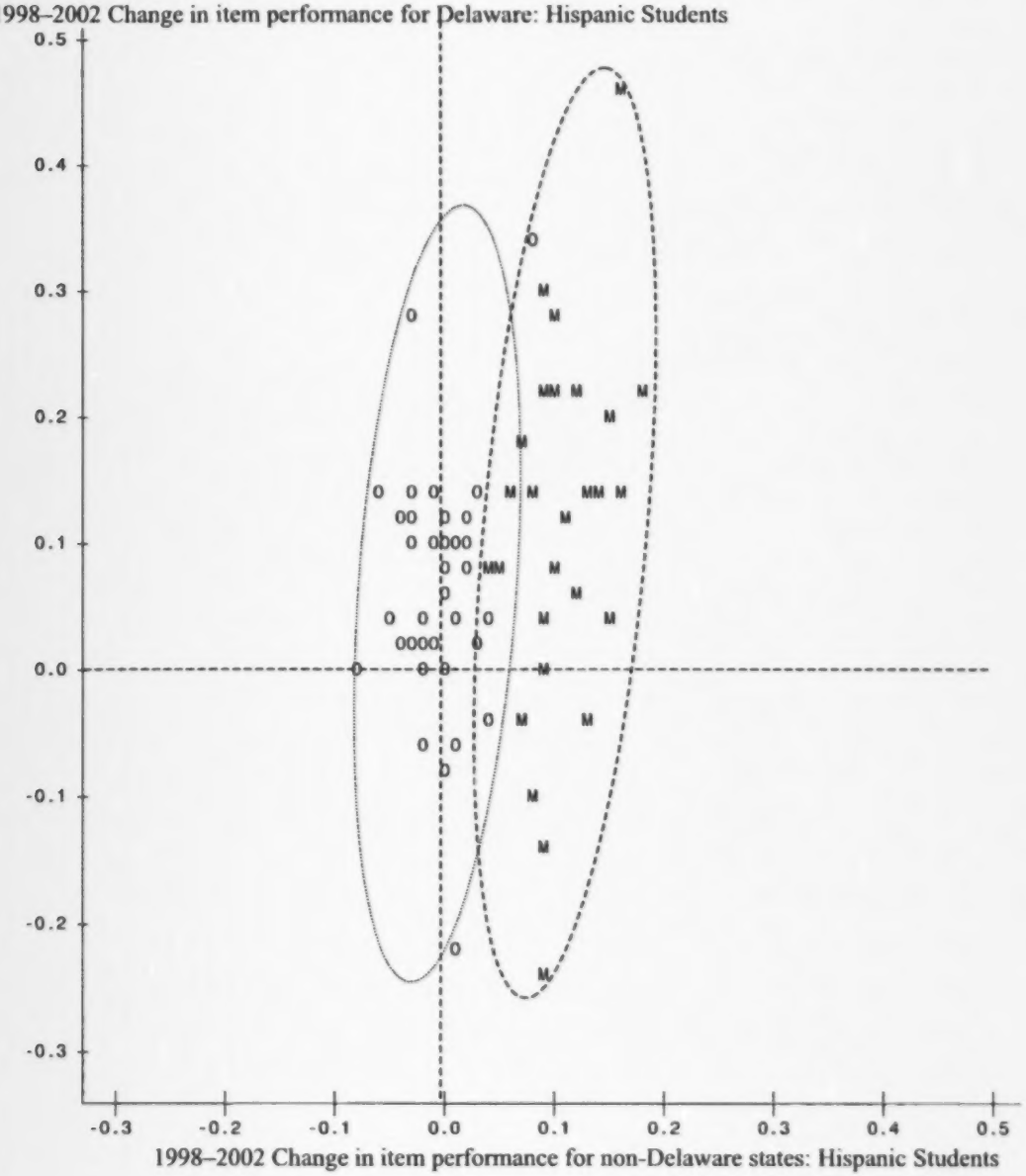
Figure 5.25. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for all students



NOTE: M=multiple choice item; O=open-ended item. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

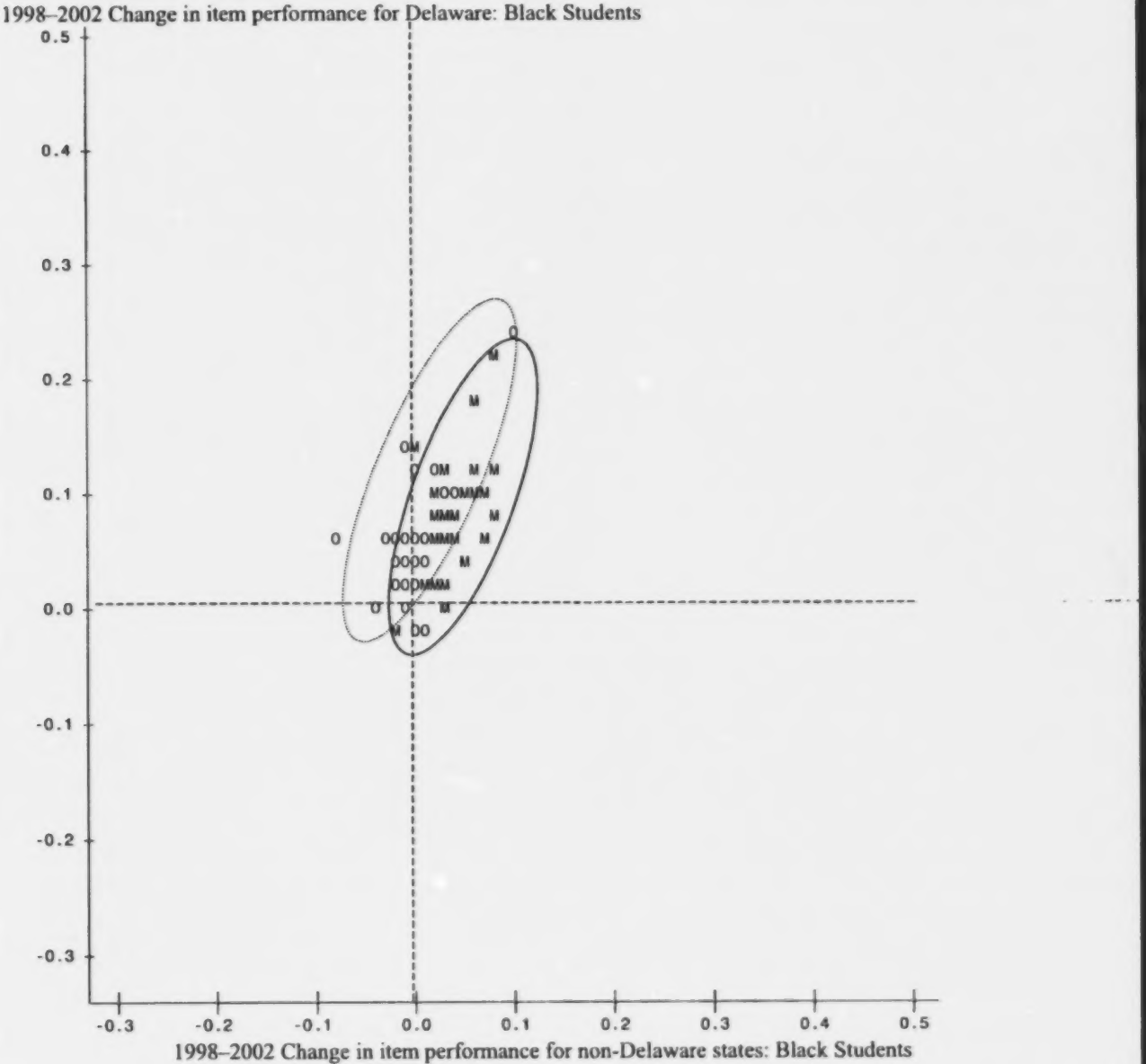
Figure 5.26. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students



NOTE: M=multiple choice item; O=open-ended item. One observation hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

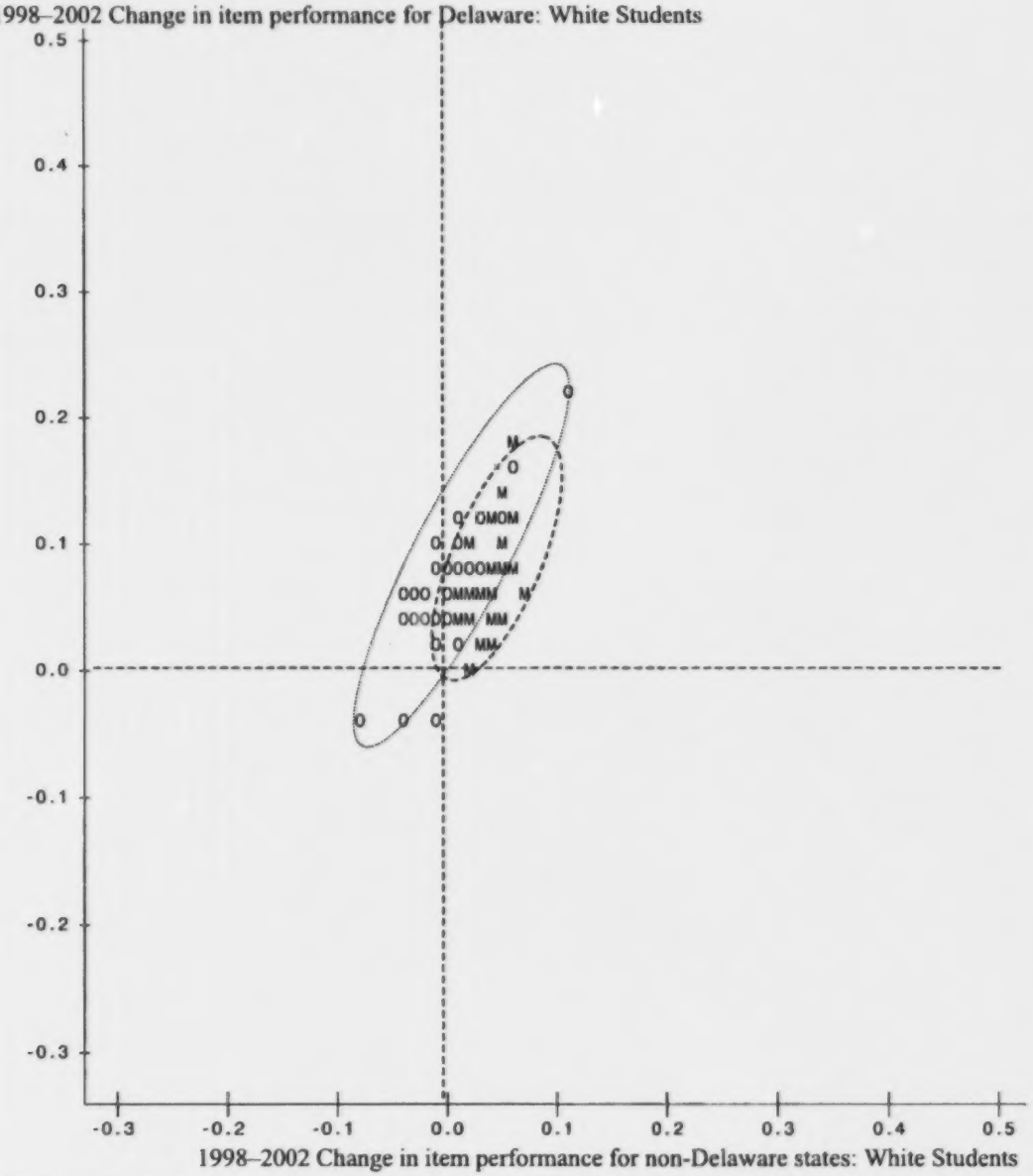
Figure 5.27. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for Black students



NOTE: M=multiple choice item; O=open-ended item. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

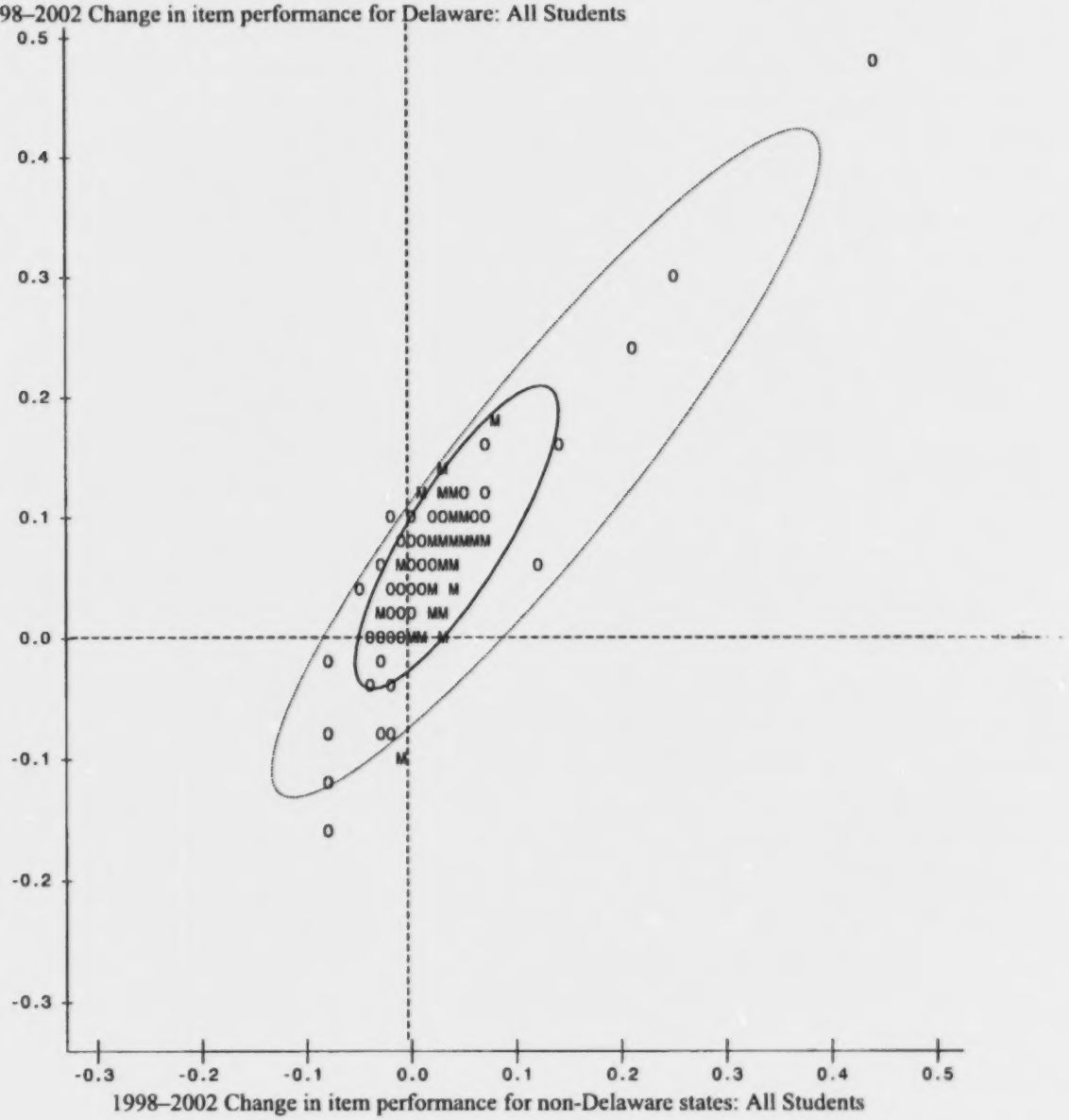
Figure 5.28. Plot of Grade 4 non-Delaware states by Delaware 1998–2002 change in item performance for White students



NOTE: M=multiple choice item; O=open-ended item. Sixteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

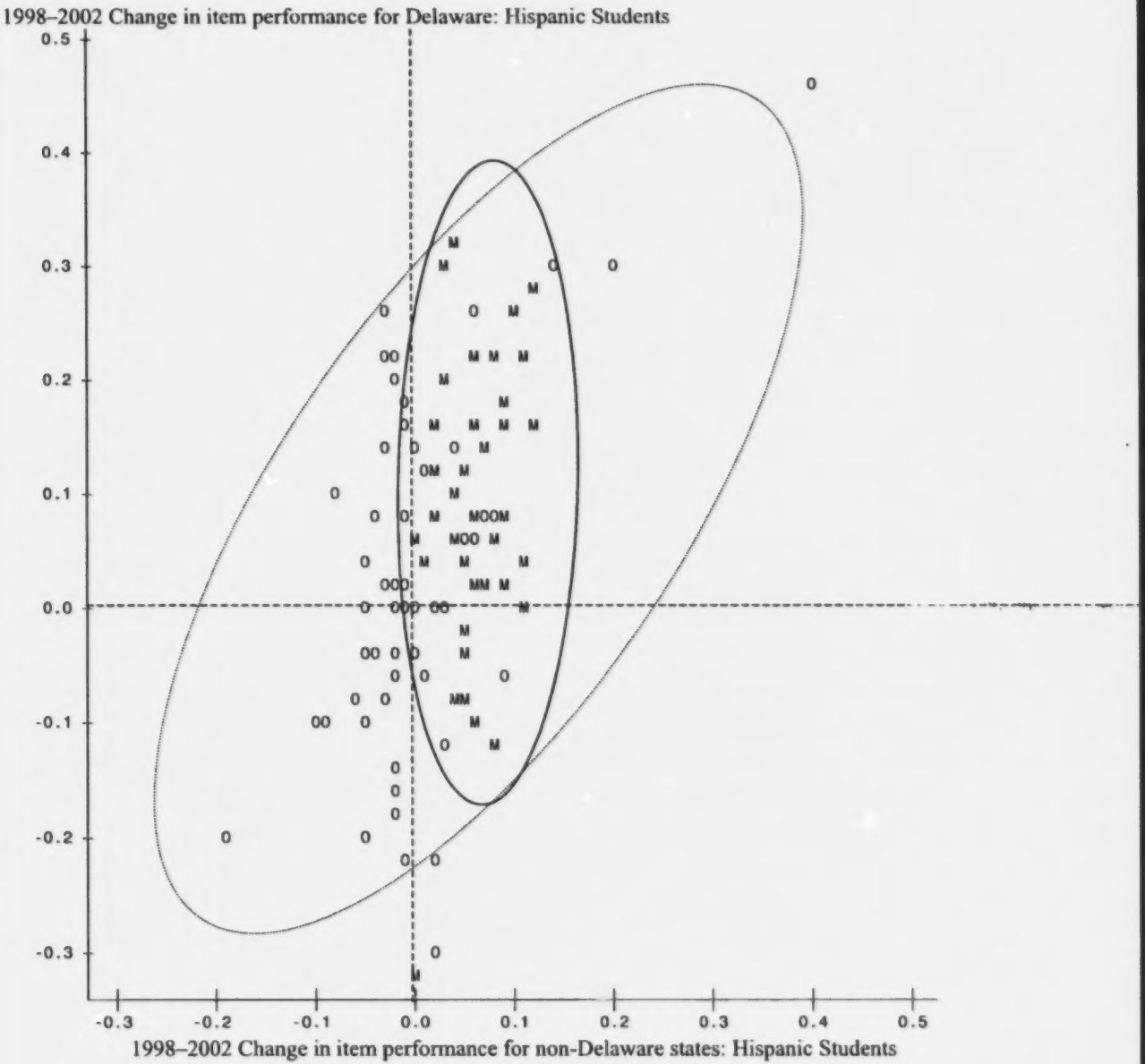
Figure 5.29. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for all students



NOTE: M=multiple choice item; O=open-ended item. Thirty-five observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

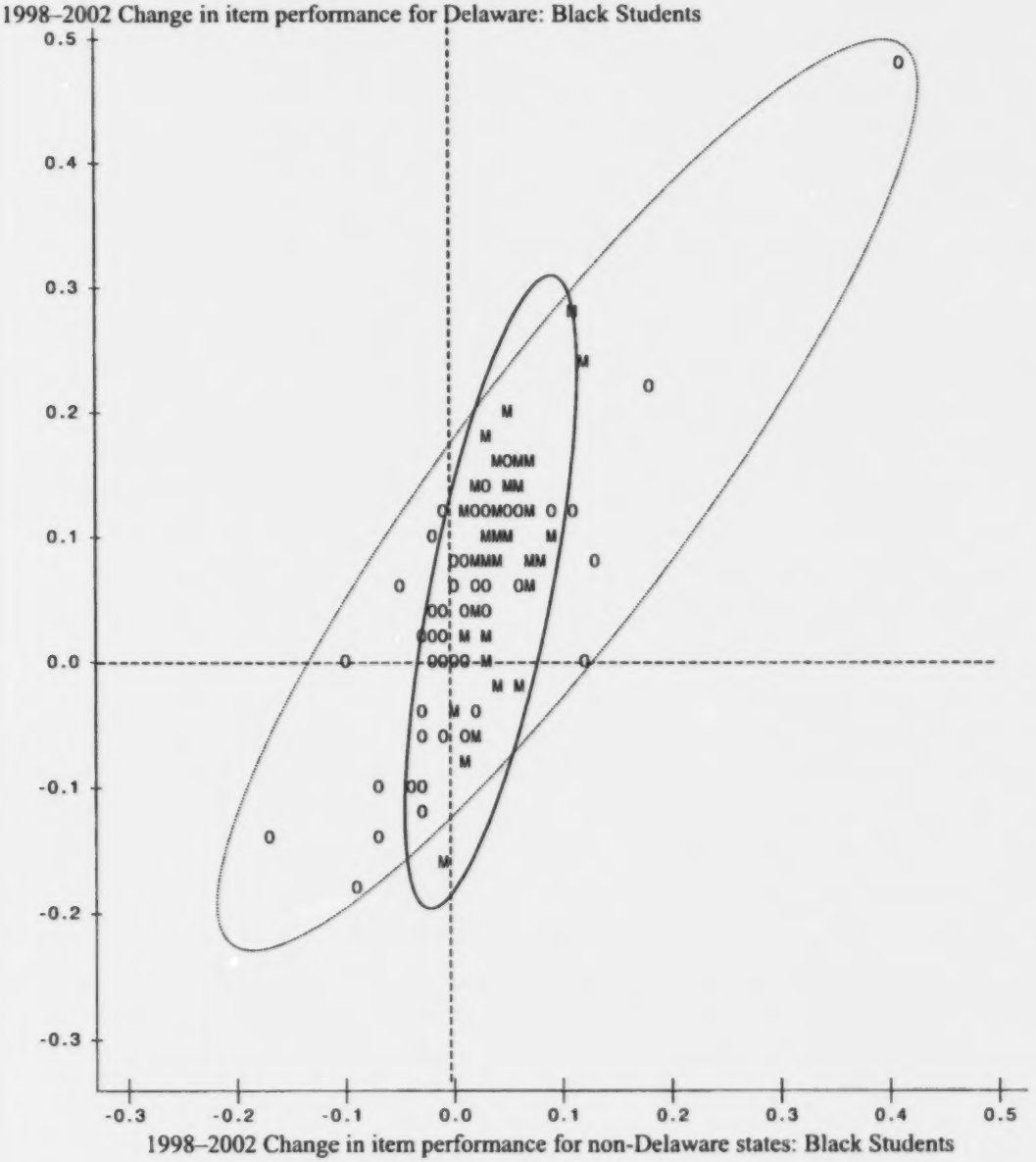
Figure 5.30. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Hispanic students



NOTE: M=multiple choice item; O=open-ended item. Thirteen observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 5.31. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for Black students.

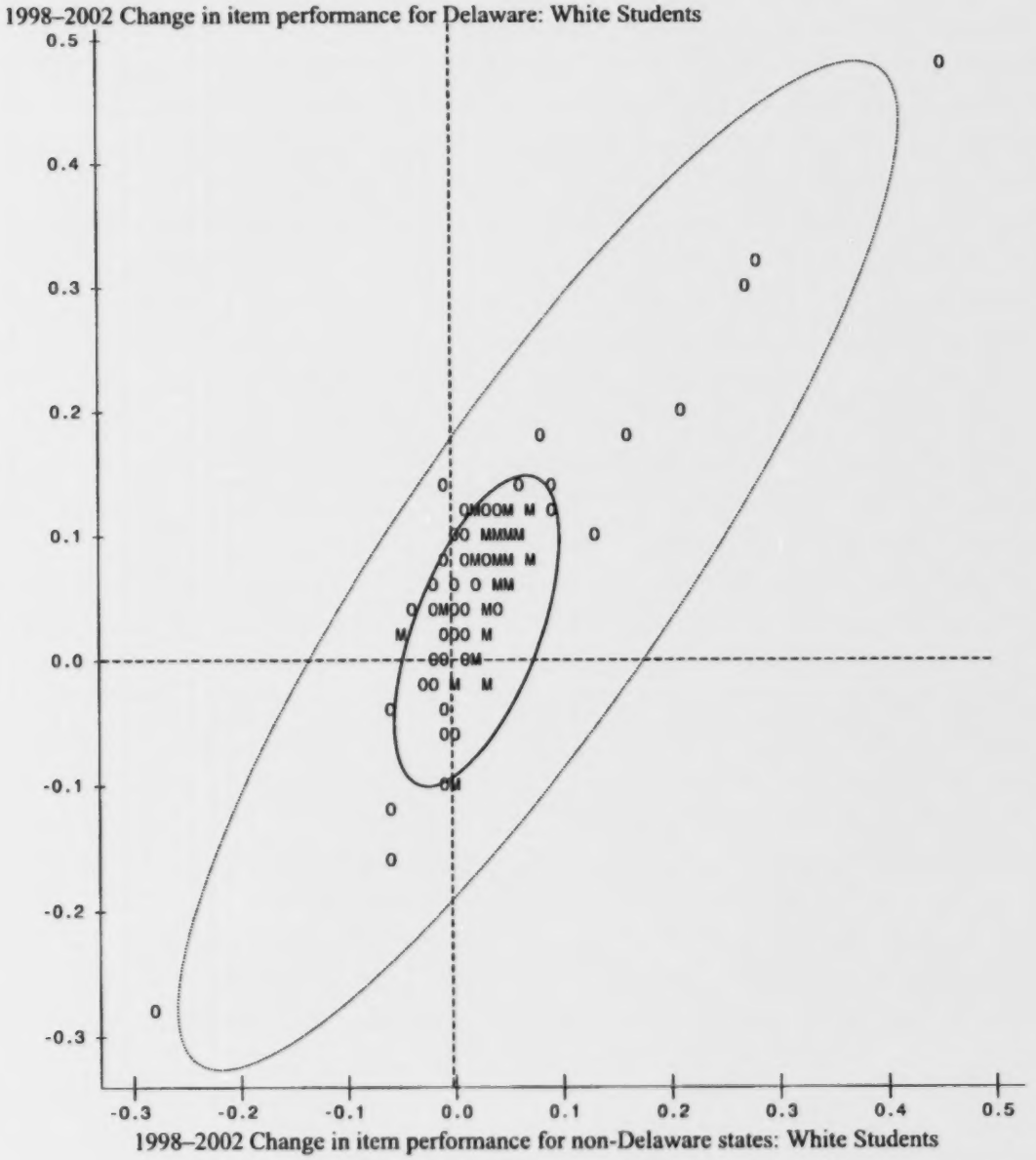


NOTE: M=multiple choice item; O=open-ended item. Twenty-four observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

BEST COPY AVAILABLE

Figure 5.32. Plot of Grade 8 non-Delaware states by Delaware 1998–2002 change in item performance for White students



NOTE: M=multiple choice item; O=open-ended item. Thirty-eight observations hidden due to overlap.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Delaware Scoring Conclusions

We found no problem in the scoring of Delaware data. Open-ended responses from Delaware students were mixed in with responses from other states in the scoring process; there was no differential treatment. Similar treatment also was found for the scanning and scoring of responses to the multiple-choice questions. Delaware students did not have unusual gains on any open-ended or multiple-choice items or passages, which might have indicated a problem with the scoring or coding, or a breach of security or exposure for any item or passage.

However, Delaware students did show slightly larger gains between 1998 and 2002 on the open-ended items relative to the rest of the nation. The improvements in open-ended items contribute to the overall gains seen by Delaware. This difference might be due to a greater emphasis on writing, which might affect success in answering open-ended items — an emphasis caused by teacher responses to the design of the state's own assessment. State contextual issues, such as state assessment configuration, are more thoroughly reviewed by one of the other investigation teams.

CHAPTER 6: SCALING AND EQUATING

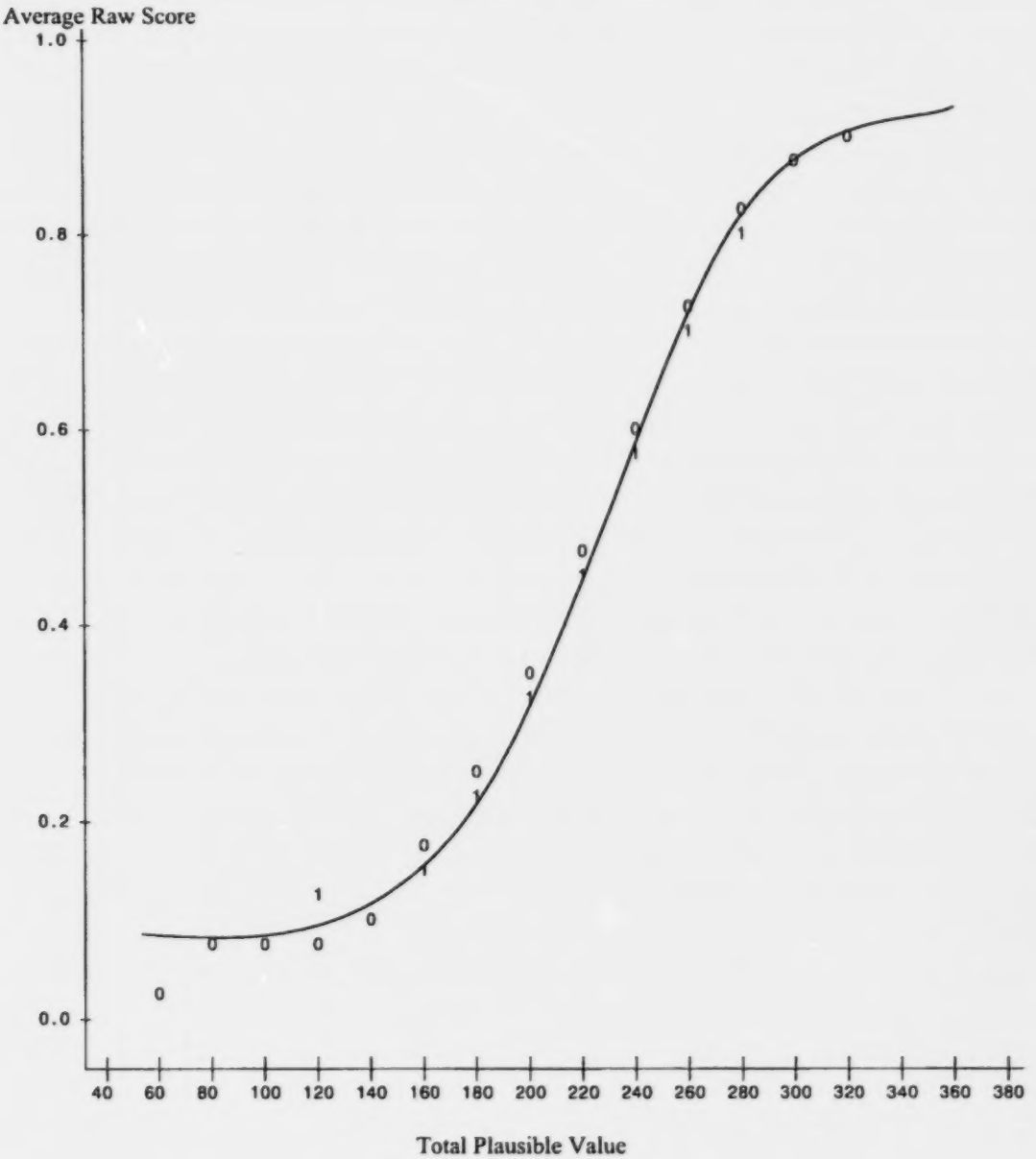
Question 5: Was scaling and equating performed correctly in Delaware?

Student scale score estimates are rooted in a sophisticated combination of item response theory (IRT) and sampling theory. IRT item parameters are used in the estimation of state score distributions and in the estimation of student plausible score values. Item parameters define the relationship between the item and the ability trait being measured. Because of the difference in results of open-ended items between Delaware and the rest of the states, one concern is whether the relationship between items and estimated ability is different for Delaware. Several analyses were conducted to look at patterns of item performance and scale scores for Delaware and the rest of the nation. In the end, a single summary figure provided a reasonable view of whether scaling and equating as applied to Delaware students was equivalent to other states.

Figure 6.1 plots the relationship between raw scores and IRT ability estimates for Grade 4 reading in 2002. Since students take different forms of the test with different numbers of possible points, HumRRO calculated raw scores for each student as the number of points earned divided by the student's possible number of points. Thus, raw scores were computed as a proportion of points earned. To avoid multiple analyses, each student's average plausible value was used as the ability estimate. The average plausible values were rounded to the nearest .20. Then, for each possible average plausible value that was represented by at least 5 students, average proportion of points earned was computed. Average proportion of points earned was computed separately for Delaware and non-Delaware states that participated in 1998. The resulting plot is essentially an IRT "test characteristic curve" captured from the data. Figure 6.1 shows non-Delaware states that participated in 1998 as a series of 0s. Only where Delaware differs does the symbol "1" appear. The curve for Delaware is essentially the same as the curves for the other states, indicating that IRT scaling and equating results must be as applicable to Delaware as to the rest of the nation. The parallel plot for Grade 8 is presented in Figure 6.2.

For another check on Delaware score means, HumRRO obtained 2001 state test data for each school in Delaware from the American Institutes for Research (AIR). HumRRO matched these data, by school, to school means that we calculated using unweighted plausible values for NAEP. Delaware's state reading test includes Grades 3, 5, and 8. Therefore, NAEP Grade 4 data were matched with state data for both Grades 3 and 5 (see Figures 6.3 and 6.4). Figure 6.5 presents the Grade 8 match. In each case, a positive relationship between NAEP scores and Delaware state test scores is highlighted by the ovals drawn on the figures. The plots add to the evidence that there were no technical errors in the processing of NAEP data for 2002.

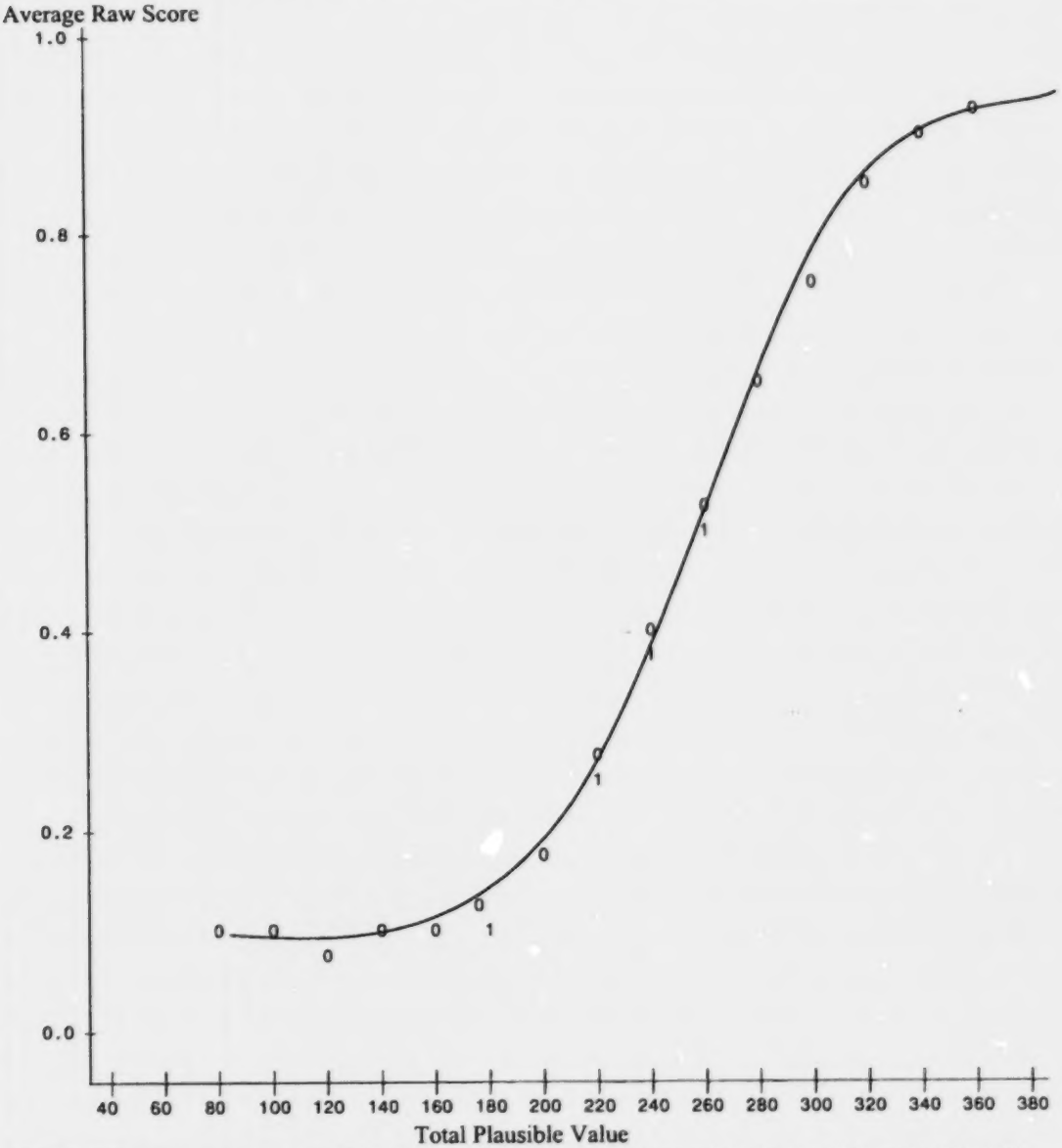
Figure 6.1. "Test characteristic curve" captured from the 2002 Grade 4 reading data



NOTE: "0" represents points on TCC for non-Delaware states that participated in 1998 and 2002. "1" identifies points where Delaware differed from other states.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

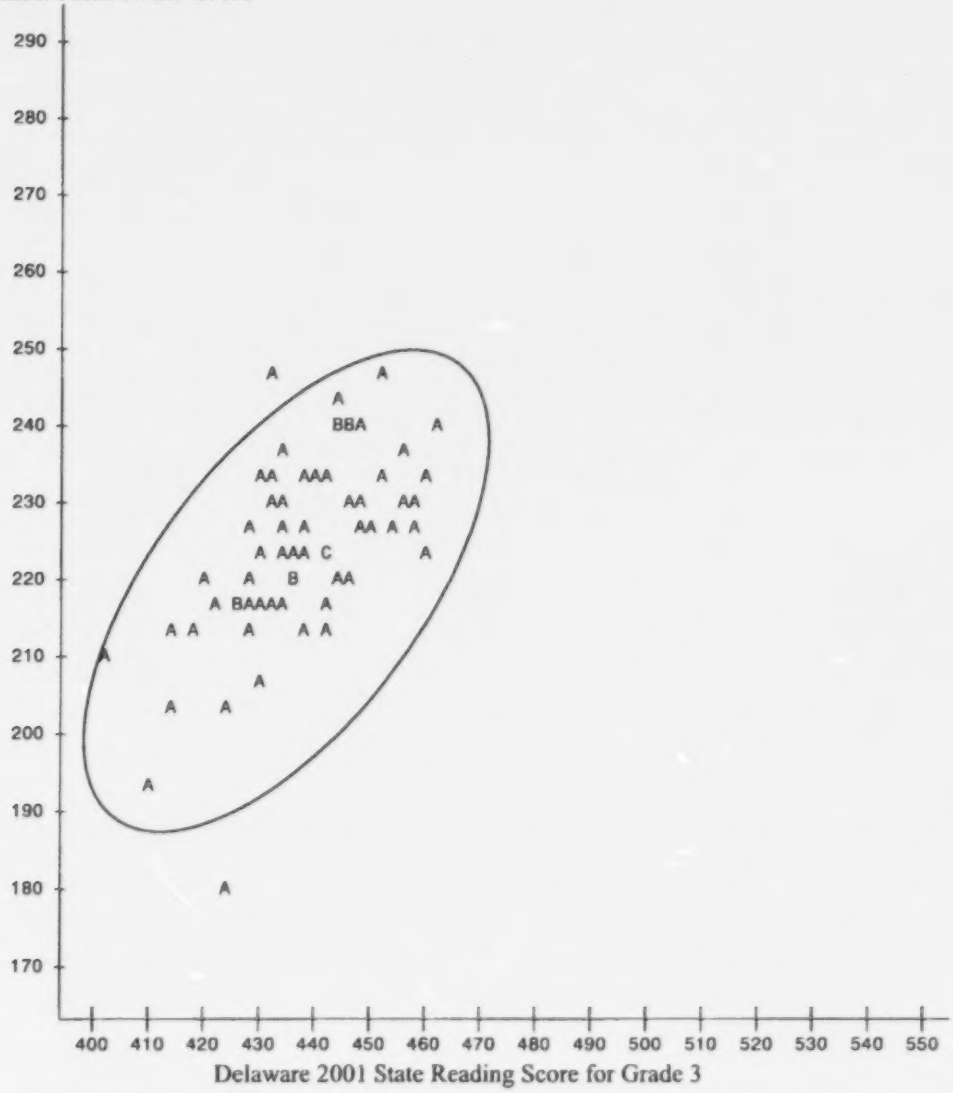
Figure 6.2. "Test characteristic curve" captured from the 2002 Grade 8 reading data



NOTE: "0" represents points on TCC for non-Delaware states that participated in 1998 and 2002. "1" identifies points where Delaware differed from other states.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

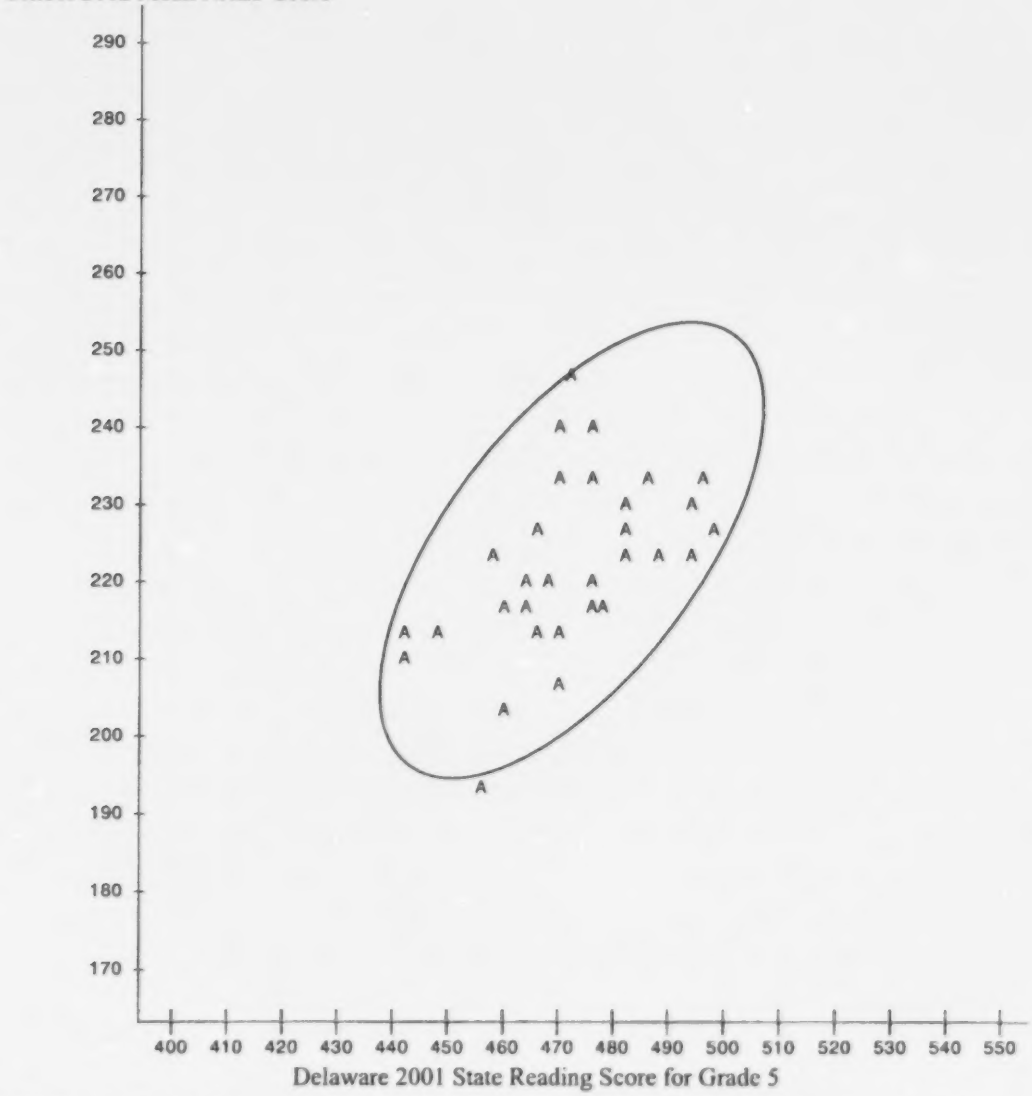
Figure 6.3. Relationship between 2001 state reading scores for Grade 3 and 2002 NAEP reading scores for Grade 4 for Delaware schools



NOTE: Letter symbol identifies number of schools represented by a single point (e.g., "A" represents one school).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

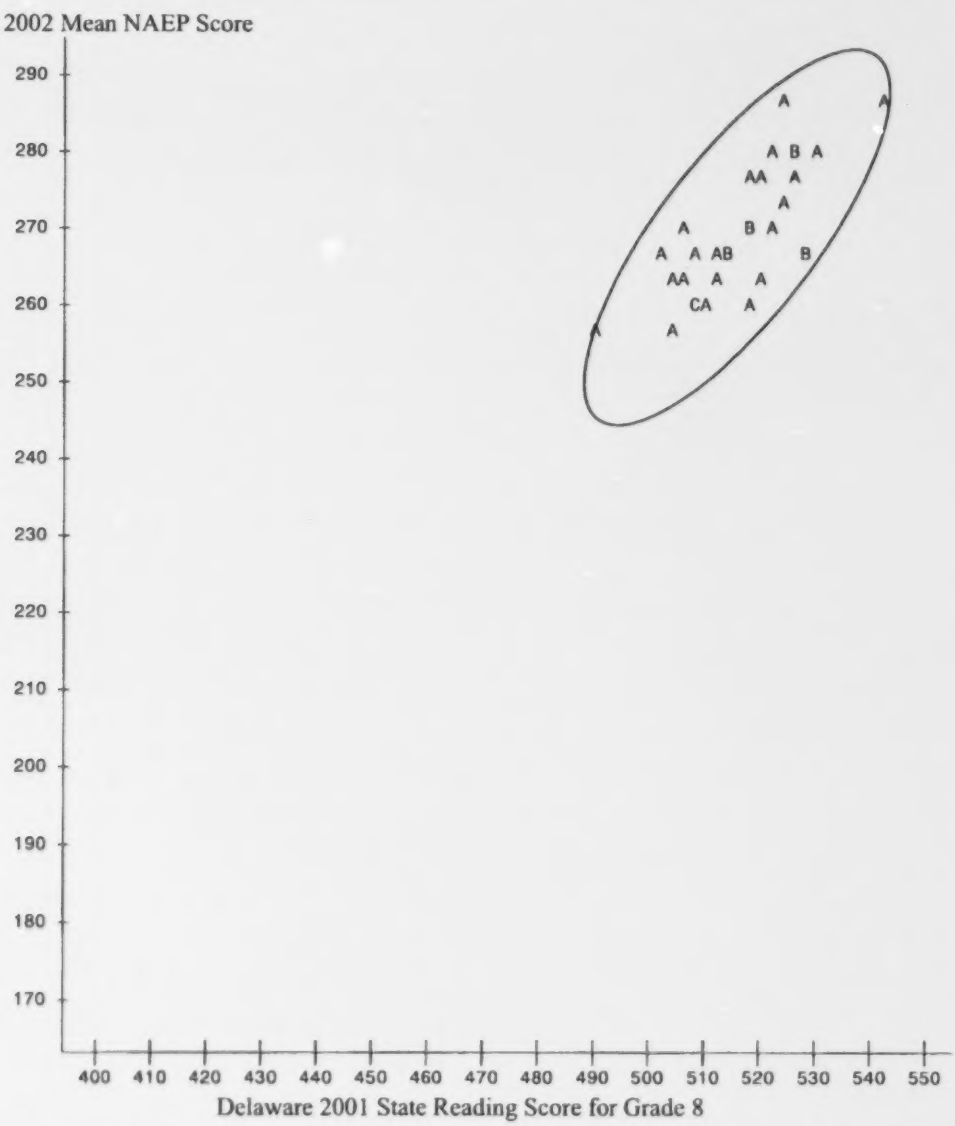
Figure 6.4. Relationship between 2001 state reading scores for Grade 5 and 2002 NAEP reading scores for Grade 4 for Delaware schools



NOTE: Letter symbol identifies number of schools represented by a single point (e.g., "A" represents one school).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 6.5. Relationship between 2001 state reading scores for Grade 8 and 2002 NAEP reading scores for Grade 8 for Delaware schools



NOTE: Letter symbol identifies number of schools represented by a single point (e.g., "A" represents one school, "B" represents two schools).

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Finally, because results for Hispanics indicated particularly large gains in the 4th grade, HumRRO independently replicated computation of 1998 and 2002 mean scores for Hispanics from plausible values and sampling weights provided in the NAEP data files. After ascertaining the appropriate data to use (e.g. school-provided race was augmented by student-reported race when school-provided race was missing), HumRRO was able to exactly reproduce the mean scores for Delaware Hispanics for 2002 produced by ETS.

Delaware Scaling and Equating Conclusions

No scaling or equating problems were identified in Delaware. We performed several analyses that examined patterns of item performance and scale scores for Delaware and the rest of the nation. As reported in the previous chapter, the relationship between scores on the individual items and scale score estimates was the same for Delaware as for other states. As shown by data presented in this chapter, Delaware results demonstrate the same relationship between scale scores and overall item mean performance as do test results for the rest of the nation. Finally, NAEP means, by school, were consistent with school means on Delaware's state test.

This page intentionally left blank.

CHAPTER 7: CODING

Question 6: Was there a problem with the coding of any data in Delaware?

The key question here is: Were background information and student responses to the questions coded correctly in the data files used for analyses? This is a particularly relevant question in light of the problem with the coding of Title 1 status in Delaware that was identified previously. The focus in this chapter was detection of gross coding errors, such as reversing codes for two demographic groups. HumRRO checked electronic filing procedures for consistency with other states and with known values, such as the proportion of students in different racial and ethnic categories.

Test results for Hispanics indicated particularly large gains in the 4th grade. This led to a specific investigation into whether the students who were coded as Hispanic were really Hispanic.

Background. Race/ethnicity results for 2002 are based on demographic information supplied by the schools, rather than on student responses to background questions. For 4th graders this was a particularly good idea given research conducted by the American Institutes for Research (AIR) showing that Grade 4 students have difficulty understanding the race/ethnicity questions. On the other hand, the schools, districts, and states supplied information electronically, through a system known as "e-filing." There had previously been a problem in Delaware with the e-filing of Title 1 information.[4] Thus, it was reasonable to ask whether the race/ethnicity information for Delaware students was correct.

Method. For the state as a whole, the distribution of race/ethnicity based on the school-report variable (SRACE) matched information from other sources reasonably well. This result did not completely answer the question, however, as it was possible that codes for two similar-sized groups might have been switched for individual schools or for the state as a whole. For the 4th grade cohort, roughly 57 percent were White, 33 percent were Black, 7 percent were Hispanic and 3 percent were Asian as reported by the schools. Switching the codes for Hispanics and Asians at some schools would lead to inaccurate score estimates for both groups.

Table 7.1 presents the relationship between student-reported race and school-reported race. Based on previous information, the pattern of relationships is as expected. Students in Grade 4 tend to over-report themselves as being Hispanic. Table 7.2 shows the relationship for Grade 8 and the expectation that Grade 8 students understand the questions, making their reports more consistent with school data.

While these data show that students and schools are not in perfect agreement, they also show that there is enough agreement to use the student data to verify whether racial/ethnic coding from the schools had (or had not) inadvertently mixed up the coding scheme. For example, if school-coded information was correctly translated into the NAEP database, then each school should show agreement rates similar to those in Tables 7.1 and 7.2.

[4] Presented by Dr. Keith Rust, Westat, at January 2003 NAEP-QA Consultant Panel Meeting.

Table 7.1. Comparison of Race/Ethnicity Codes from Schools and Students in the 2002 Delaware 4th Grade Reading Assessment

School Report		Percent in Each Student Report Category				
Race/Ethnicity	N	White	Black	Hispanic	Asian	Amer. Ind.
White	2,424	88.7%	1.2%	5.5%	1.2%	3.5%
Black	1,383	1.1%	82.4%	11.5%	0.7%	4.4%
Hispanic	291	3.1%	0.7%	95.9%	0.3%	0.0%
Asian	228	6.3%	0.8%	10.9%	78.9%	3.1%
Amer. Ind.	17	47.1%	11.8%	5.9%	0.0%	35.3%

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 7.2. Comparison of Race/Ethnicity Codes from Schools and Students in the 2002 Delaware 8th Grade Reading Assessment

School Report		Percent in Each Student Report Category				
Race/Ethnicity	N	White	Black	Hispanic	Asian	Amer. Ind.
White	2,573	92.1%	1.4%	3.5%	1.0%	1.9%
Black	1,210	0.8%	86.8%	7.6%	1.2%	3.6%
Hispanic	240	0.4%	0.4%	98.8%	0.4%	0.0%
Asian	100	0.0%	1.0%	7.0%	91.0%	1.0%
Amer. Ind.	19	21.1%	10.5%	26.3%	0.0%	42.1%

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

For each school, HumRRO looked at each racial/ethnic category reported by the school. The frequency with which students placed themselves in a category was compared to the frequency with which students in the school-based grouping reported themselves in each category. Every instance where another category was more frequently selected for any of the school report groups was flagged. In all, 71 schools had non-discrepant results and 15 schools were flagged as having potentially discrepant results, in all cases for only one racial category. Table 7.3 summarizes discrepancies found for these 15 cases. In 12 cases, a single student who was the only student in a school category caused the discrepancy and who reported a different category, that is, two of the other three cases did involve students coded by the school as Hispanic. In all, the number of discrepancies was very small and each discrepancy involved only one or two students.

For 8th grade students, race/ethnicity code agreement at the school level was higher than for 4th grade students. As shown in Table 7.4, there were only seven schools for which there was a school race category in which students selected some other category more frequently. Again, all instances involved only a very small number of students and, in this case, all of the differences involved the American Indian category for which sample sizes were too small to support reporting.

Table 7.3. Discrepancies between School and Student Race/Ethnicity Codes for Individual Delaware Schools in the NAEP 2002 4th Grade Reading Assessment

Cases with Only a Single Student in the School Race Category		
School Race	Student Race	Occurrences
Hispanic	Asian	2
American Indian	White	5
American Indian	Hispanic	1
Asian	White	1
Hispanic	White	1
Black	White	1
White	American Indian	1

Cases with More than One Student in the School Race Category			
School Race	N	Most Frequent Student Race	Next Most Frequent Student Race
Hispanic	5	Asian (3)	Hispanic (2)
Hispanic	3	White (2)	Black (1)
American Indian	3	White (2)	Black (1)

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Table 7.4. Discrepancies between School and Student Race/Ethnicity Codes for Individual Delaware Schools in the NAEP 2002 8th Grade Reading Assessment

Cases with Only a Single Student in the School Race Category		
School Race	Student Race	Occurrences
American Indian	White	1
American Indian	Hispanic	1
American Indian	Black	1
Asian	American Indian	1

Cases with More than One Student in the School Race Category			
School Race	N	Most Frequent Student Race	Next Most Frequent Student Race
American Indian	2	Hispanic (2)	N/A
American Indian	3	Hispanic (2)	White (1)
American Indian	3	White (2)	American Indian (1)

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Delaware Coding Conclusions

No coding problems were found. Racial/ethnic codes used for reporting were reviewed because of large gains for one category of students. Agreement between race/ethnicity data supplied by students and by schools was sufficient to rule out coding errors, overall and for each school.

This page intentionally left blank.

CHAPTER 8: TEST SECURITY

Question 7: Was there a breach in test security in Delaware?

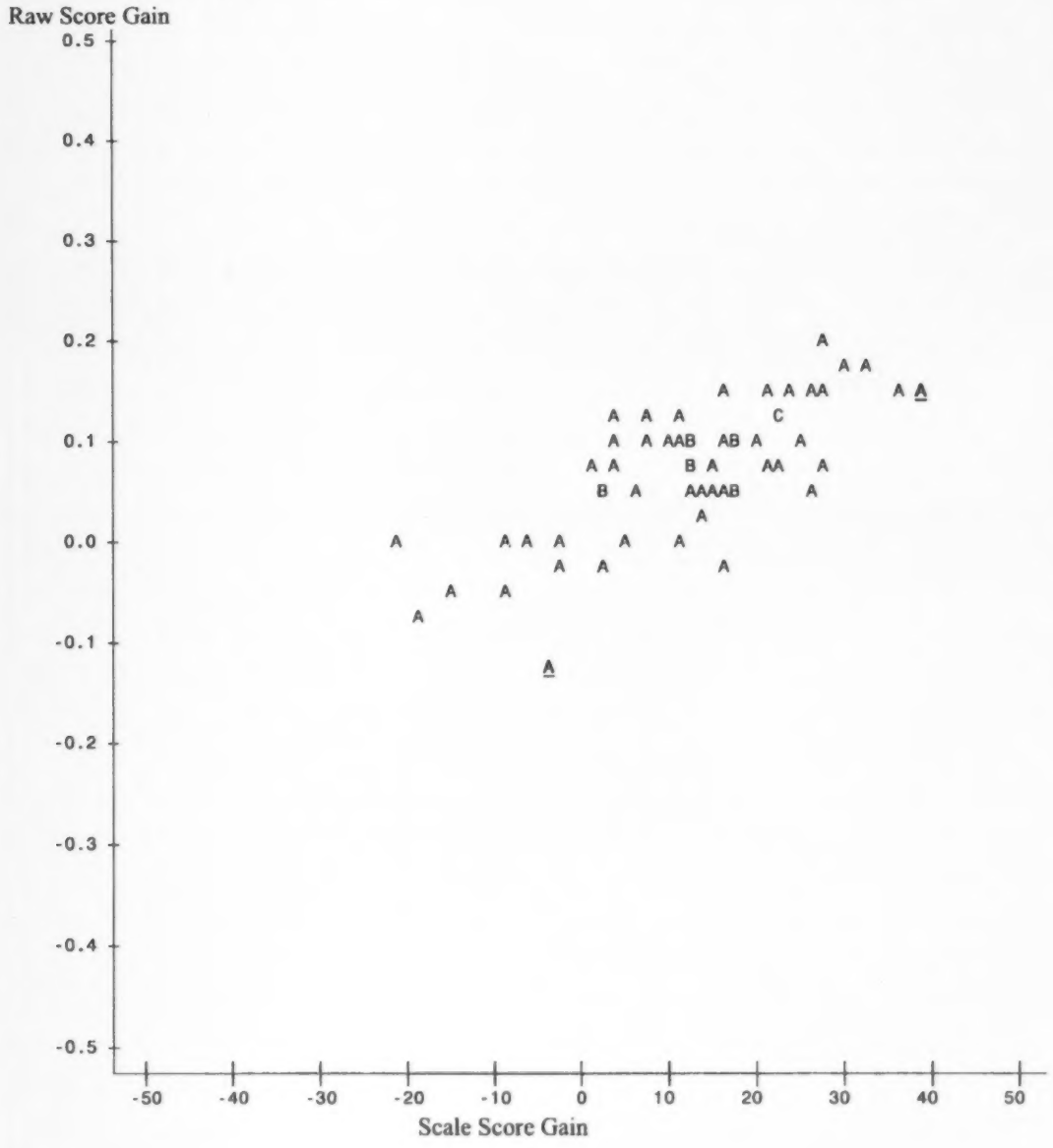
Chapter 5 has already presented comparisons between Delaware and the rest of the nation for item-level performance and shown no patterns that suggest statewide breach of security. In this chapter, we investigated the question by looking for schools whose 2002 data were inconsistent with their 1998 data. To test this, we created scatter-plot diagrams of school-level gains on item points versus gains on scale scores. The plots provide two ways (via scale score and raw scores) for identifying schools with particularly high gains. For any suspect school, the gains on each item were examined, looking for unusual gains on items associated with a common reading passage.

The matching of schools across years was somewhat surprising, particularly for Grade 4. Although all Delaware schools were tested in both 1998 and 2002, 65 Grade 4 schools were tested in 1998 and 86 in 2002, a net increase of 21 schools. Fifty-nine of these schools were positively matched using NCES school codes, but several other 1998 schools may not have been matched, due to changes in their codes. Because of the differences, HumRRO conducted an accounting of the 1998 schools. Delaware reported 87 schools in 1998, 22 more than the 65 in the 1998 NAEP sample. A by-name list of these schools revealed them to be special schools, the majority of which were "Intensive Learning Centers," which would not be sampled by NAEP. Gains for the 59 matched Grade 4 schools appear in Figure 8.1.

To help identify "unusual" schools, a parallel scatter plot was constructed for the 805 schools outside of Delaware that participated in 1998 as well as in 2002. The range of gains and losses for Delaware (Figure 8.1) and non-Delaware schools (Figure 8.2) is similar. Gains as high as 30 and 40 scale score points were not uncommon among the non-Delaware schools. Of course, item exposure could exist in any of these Delaware or non-Delaware schools. The data show that Delaware gains are within the range of gains for the rest of the nation.

The item-level changes for the two schools highlighted (bold and underlined) in Figure 8.1 were also examined closely. The top right school showed some high gains (greater than .3) for items in two passages. The lower, center school, was examined because it was well outside of the pattern of the other schools. This school showed both large gains and large losses on various items throughout the test. The item-level gain data for the remaining schools were also scanned, but did not reveal any suspect patterns (i.e., gains on a passage that were high and/or discrepant from the rest of the item-level gains of the school).

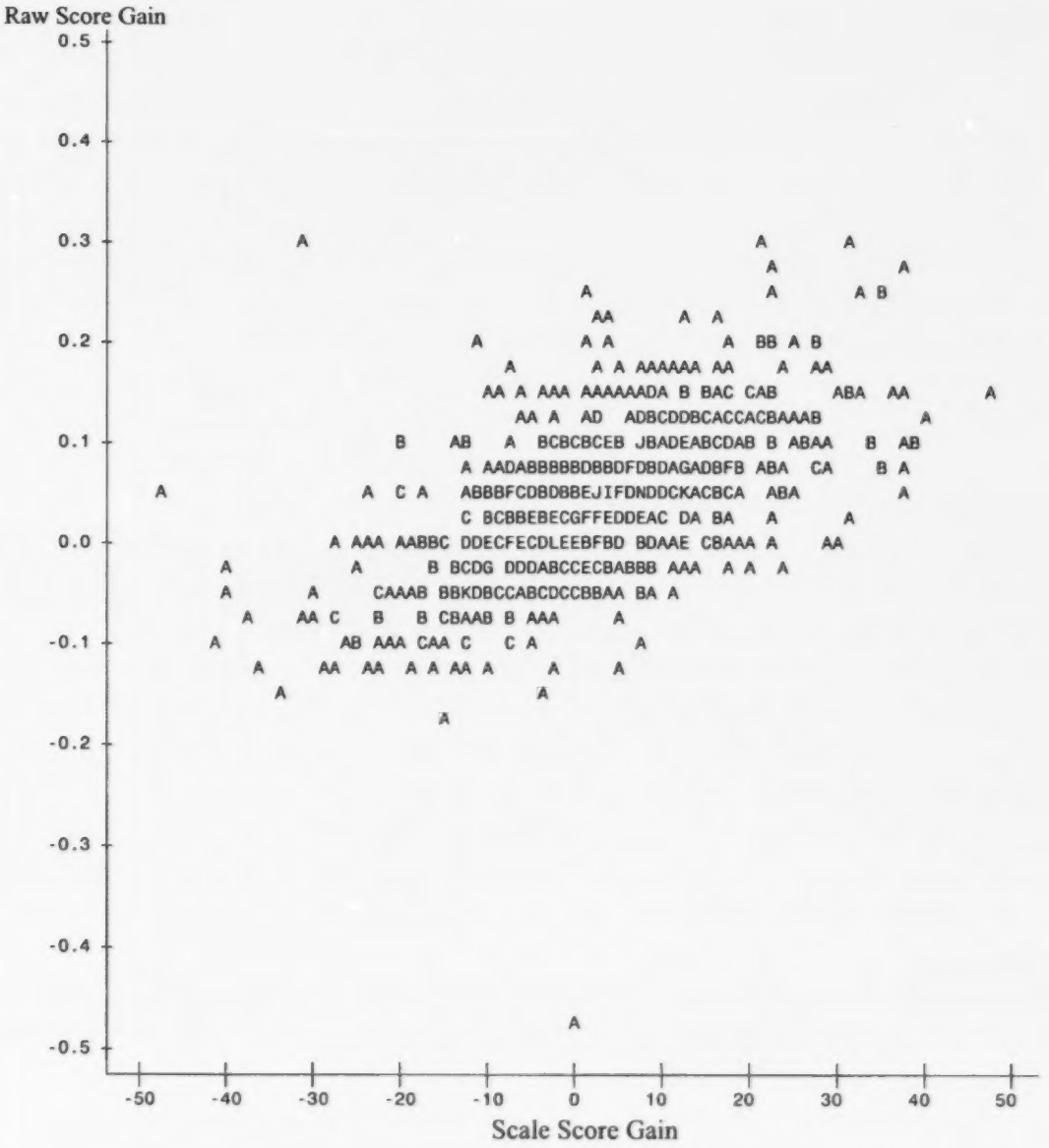
Figure 8.1. Delaware 1998–2002 4th grade school gains on raw score versus gain on NAEP scale scores



NOTE: A = 1 observation; B = 2 observations, etc.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 8.2. Non-Delaware 1998–2002 4th grade school gains on raw score versus gain on NAEP scale scores



NOTE: A = 1 observation; B = 2 observations, etc.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

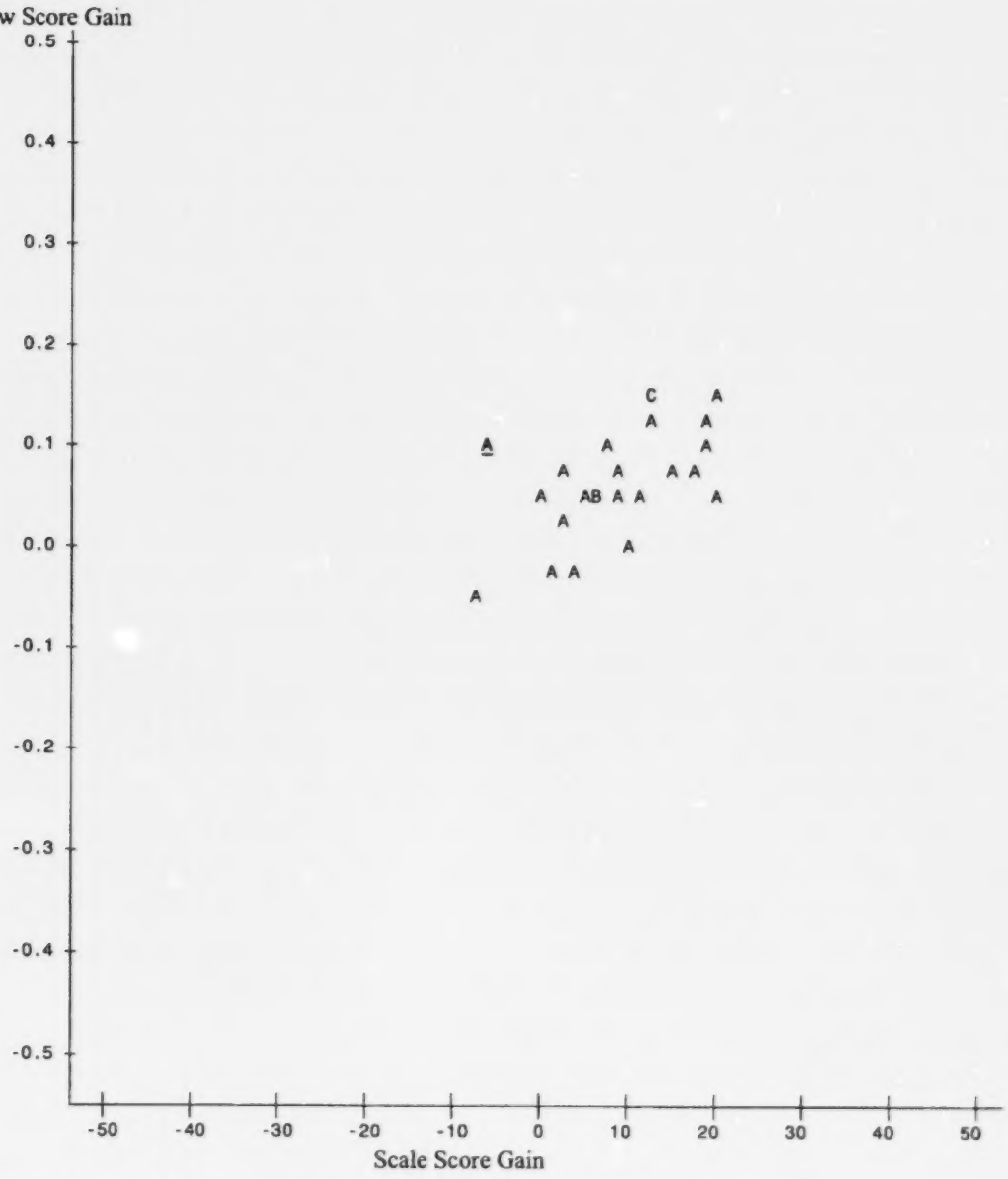
Figures 8.3 and 8.4 repeat the by-school gain analysis for Grade 8 schools. Of the 28 Grade 8 schools tested in 1998 and the 35 tested in 2002, 25 schools were matched. The pattern is more scattered, and the highest gains are not as large as those for Grade 4. On the other hand, the gains for Delaware schools all fall within the range of the gains for the rest of the nation.

The one school (bold and underlined in Figure 8.3) with a loss in scale score and a gain in raw score showed, in general, losses and small gains on most items, but had large gains (.39 to .55) on about a half dozen items clustered in two or three passages. However, none of the passages showed gains on a majority of items in the passage. While there were passages that contained two or three items with large gains, those passages also had items with small gains and losses in performance.

Delaware Security Conclusions

No indications of test security breaches were identified. Gains on individual items and on blocks of items associated with a common passage were consistent with gains on these items and blocks for the nation as a whole. For the few individual schools that did not show unusual gain patterns overall, there was no consistent pattern to item-level gains.

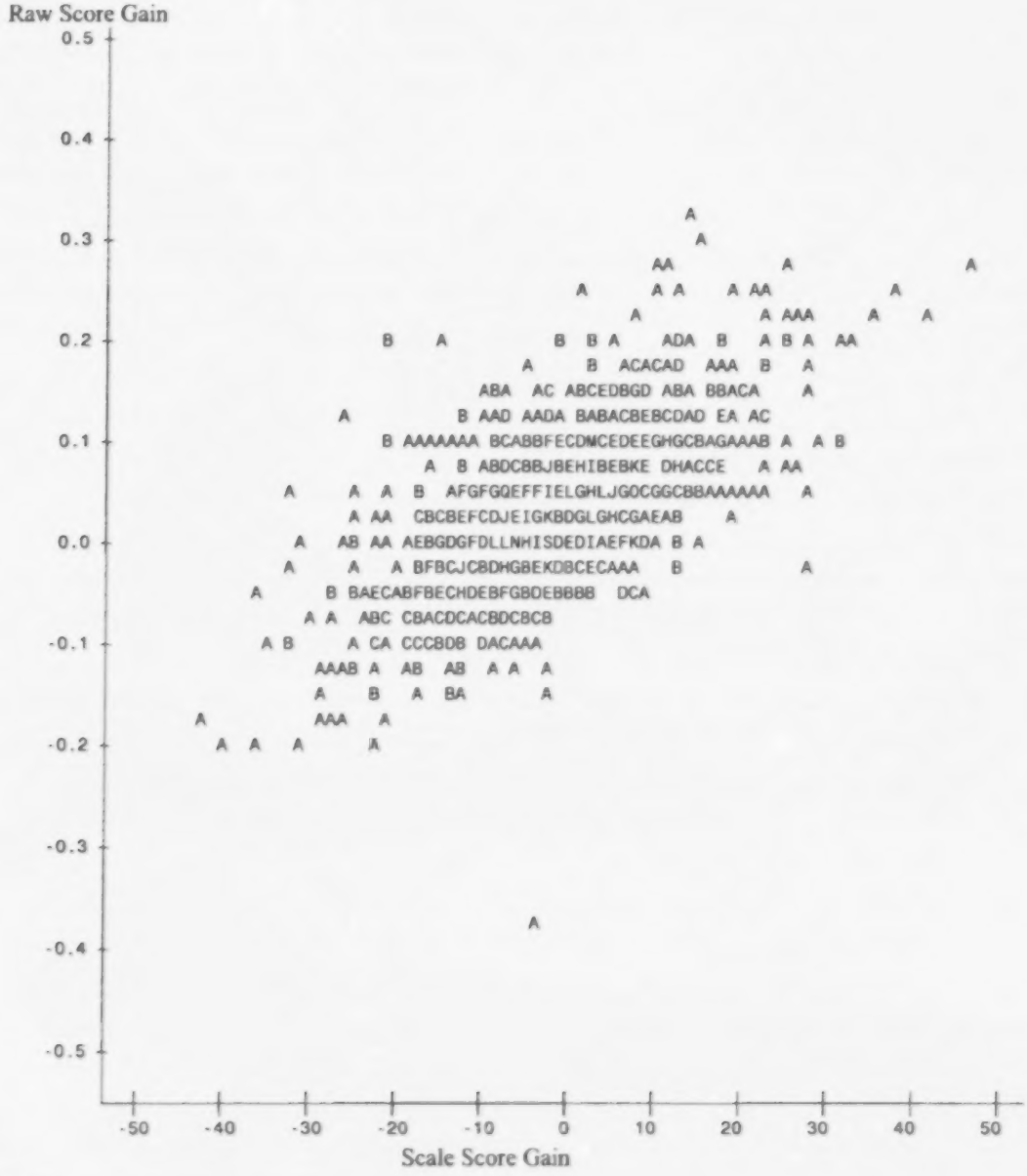
Figure 8.3. Delaware 1998–2002 8th grade school gains on raw score versus gain on NAEP scale scores



Legend: A = 1 observation; B = 2 observations, etc.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

Figure 8.4. Non-Delaware 1998–2002 8th grade school gains on raw scores versus gain on NAEP scale scores



NOTE: A = 1 observation; B = 2 observations, etc.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1998 and 2002 Reading Assessments.

CHAPTER 9: CONCLUSIONS

HumRRO investigated the seven specific questions that were identified by NCES. Our findings supported the sampling, weighting, BIB spiral, scoring, scaling and equating, and test security conclusions drawn by NAEP Alliance contractors. However, we detected one related problem that would justify caution in interpretation of the 2002 estimates of Hispanic gains.

Prior to calculation of the gains between 2002 and 1998, the 1998 results were recomputed with two changes:

- Contractors used an alternate sample of students, who were provided accommodations similar to those provided in 2002, in making the recomputation.
- Contractors defined race categories from codes supplied by schools rather than by students. Consequences of these changes affected the sample size, mean, standard error of the mean, and exclusion rate.

Among Delaware fourth graders, the recomputation lowered the 1998 Hispanic mean from 193 to 176. Sample size for Hispanics decreased from 198 to 101. The standard error of the Hispanic mean increased from 4 to 12 scale points. The 1998 exclusion rate for Hispanics dropped from 6 percent to 3 percent. Among Delaware eighth graders, the recomputation raised the 1998 Hispanic mean from 246 to 248. Sample size for Hispanics decreased from 78 to 64. The standard error of the Hispanic mean decreased slightly from 9 to 8 scale points. The 1998 exclusion rate for Hispanics dropped from 12 percent to 0 percent.

We recommend that the Delaware Hispanic gains from 1998 to 2002 be flagged in some way to indicate that the amount of gain may be distorted by small sample size, high standard errors, and large changes in exclusion rates.

In summary, based on an extensive analysis of 2002 Delaware reading assessment data and on data from the 1998 assessment used as the basis for computing gains in 2002, we did not find any methodological or technical procedural problems that could have affected the 2002 results for Delaware. We did note that revised 1998 score estimates for 4th grade Hispanic students had large standard errors. We recommend that estimated gains for 4th grade Hispanic students computed from these revised estimates be flagged with appropriate explanatory text.

This page intentionally left blank.

APPENDIX A: OUTSIDE REVIEW OF NAEP SAMPLING PROCEDURES

Chapter 2 indicates that Delaware is unique in two respects regarding the selection of students for NAEP. First, all eligible schools are included for both 1998 and 2002, making Delaware different from the remaining states for which schools are sampled. In addition, the same contrast between full census and sampling occurred within Delaware with regard to the selection of students within schools. In 1998, students were sampled. In 2002, all students were tested. The effects of this difference in selection methodology—sampling a given population or attempting to test all of that same population—are addressed below.

Samples versus Censuses

Consider whether a sample yields a different result from a census. If sampling is considered part of a continuum up to and including a census as a 100 percent sample, then there is no reason to believe that a sample would yield a result in any way different than a census. If the operations conducted to collect the information are exactly the same, regardless of the size of the sample (up to and including 100 percent) then the values obtained from individual respondents would be the same regardless of how large the sample is.

The only issue that remains is how the results from the sample are projected to the population. The simplest case is a census: each response represents only itself and no other response from the population being studied. The responses are simply aggregated and averaged.

The next simplest case is a simple random sample. Suppose that only half of the students in an area are surveyed to measure educational progress. Each student in the sample represents one other student in the population who was not contacted. However, the process is the same in terms of using the information from the survey. The answers are aggregated and averaged. The average from the sample represents the best estimate of the average in the population. There is no reason to believe that this process would yield a value different from the population value being estimated, except by chance. And in this process, we are equally likely to be slightly high or slightly low in estimating the average educational progress for the group. Typically, the larger the sample, the smaller the chance (or error) variance in population values.

We can make these scenarios increasingly complex. But for each level of complexity added, the process of projection from the sample to the population is essentially the same. We add the results from the sample and average them. If some groups in the population have greater proportional representation than others in the population, we weight the results together so that contributions to the overall average are in proportion to their proper weight in the population.

Are there ways in which sample results might differ from census results? Yes – it is possible for sample results to differ from census results for reasons that are not statistical. Attempting to contact all of the population can be a relatively expensive undertaking. If the researcher is not cautious in how expenditures are made, the quality of data in a census may deteriorate relative to the data that could come from a sample. For a fixed budget, if more resources are channeled into frenetically contacting schools and students and fewer resources are available to collect the data, then the quality of the data may suffer in the census. If a proper balance is maintained in contacting schools and students and the collection of data from these

sources, then there is no reason to believe that the results would differ whether 50 percent, 80 percent, or 100 percent of the students are interviewed.

Sample and census results may differ for one other reason, related to the previous discussion. Those people who are most difficult to find and contact are sometimes different in their characteristics from the balance of the population. For example, students with the worst attendance records will be the students most difficult to contact. These students are also likely to be the ones to show the least progress when measuring educational attainment. If the survey only contacts students on one day and no attempts are made to follow up students who were absent, then an upward bias might result in the survey. A strong follow-up program would alleviate this problem, as the resources expended for the follow-up in the sample would be similar to the types of resources necessary to complete a census. There would be a proportional representation of students who are likely to complete school and those likely to drop out.

The same argument can be made at the local education agency (LEA) and school level. A census of these would naturally make every effort to include every LEA in a state, or every school in an LEA. If a sample is selected, the same efforts need to be made to include the sampled LEAs or the sampled schools. With a properly designed sample, LEAs or schools that would decline to be in the census would be proportionally represented in the sample. Put another way, if 10 percent of the schools in a census of schools would decline to participate, we would expect that on average 10 percent of the schools selected for the sample would decline to participate. If the same resources are put into converting refusals in the sample or the census, there is no reason to believe that the census would be any better or any worse than the sample.

Summary

There is no reason to believe that using a sample in any way produces a result that would be different from the result that would be obtained by conducting a census. The only difference between the sample estimate and the result from the census is that there is some uncertainty associated with the sample estimate and how close it may be to the true value being estimated. With a sufficiently large sample, proper design, and the same efforts at execution, this variance from the true value will be negligible and not material to any decision-making process using the survey results. Conversely, if sample sizes are too small, error variance may increase so much that the data are not useable.

Listing of NCES Working Papers to Date

Working papers can be downloaded as PDF files from the NCES Electronic Catalog (http://nces.ed.gov/pubsearch/). You can also contact Sheilah Jupiter at (202) 502–7363 (sheilah.jupiter@ed.gov) if you are interested in any of the following papers.

Listing of NCES Working Papers by Program Area

No.	Title	NCES contact
Baccalaureate and Beyond (B&B)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Beginning Postsecondary Students (BPS) Longitudinal Study		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Common Core of Data (CCD)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–12	Coverage Evaluation of the 1994–95 Common Core of Data: Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2002–02	School Locale Codes 1987 - 2000	Frank Johnson
Data Development		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Decennial Census School District Project		
95–12	Rural Education Data User's Guide	Samuel Peng
96–04	Census Mapping Project/School District Data Book	Tai Phan
98–07	Decennial Census School District Project Planning Report	Tai Phan
Early Childhood Longitudinal Study (ECLS)		
96–08	How Accurate are Teacher Judgments of Students' Academic Performance?	Jerry West
96–18	Assessment of Social Competence, Adaptive Behaviors, and Approaches to Learning with Young Children	Jerry West
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle Childhood	Elvira Hausken

BEST COPY AVAILABLE

No.	Title	NCES contact
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
Education Finance Statistics Center (EDFIN)		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
Education Longitudinal Study: 2002 (ELS:2002)		
2003-03	Education Longitudinal Study: 2002 (ELS: 2002) Field Test Report	Jeffrey Owings
High School and Beyond (HS&B)		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	**Dawn Nelson**
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
HS Transcript Studies		
1999–05	Procedures Guide for Transcript Studies	**Dawn Nelson**
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
International Adult Literacy Survey (IALS)		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Integrated Postsecondary Education Data System (IPEDS)		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
National Assessment of Adult Literacy (NAAL)		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom

BEST COPY AVAILABLE

No.	Title	NCES contact
National Assessment of Educational Progress (NAEP)		
95–12	Rural Education Data User's Guide	Samuel Peng
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Steven Gorman
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Steven Gorman
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Steven Gorman
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questionnaires)	Steven Gorman
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Steven Gorman
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-06	NAEP Validity Studies: The Validity of Oral Accommodation in Testing	Patricia Dabbs
2003-07	NAEP Validity Studies: An Agenda for NAEP Validity Research	Patricia Dabbs
2003-08	NAEP Validity Studies: Improving the Information Value of Performance Items in Large Scale Assessments	Patricia Dabbs
2003-09	NAEP Validity Studies: Optimizing State NAEP: Issues and Possible Improvements	Patricia Dabbs
2003-10	A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments	Marilyn Binkley
2003-11	NAEP Validity Studies: Reporting the Results of the National Assessment of Educational Progress	Patricia Dabbs
2003-12	NAEP Validity Studies: An Investigation of Why Students Do Not Respond to Questions	Patricia Dabbs
2003-13	NAEP Validity Studies: A Study of Equating in NAEP	Patricia Dabbs
2003-14	NAEP Validity Studies: Feasibility Studies of Two-Stage Testing in Large-Scale Educational Assessment: Implications for NAEP	Patricia Dabbs
2003-15	NAEP Validity Studies: Computer Use and Its Relation to Academic Achievement in Mathematics, Reading, and Writing	Patricia Dabbs
2003-16	NAEP Validity Studies: Implications of Electronic Technology for the NAEP Assessment	Patricia Dabbs
2003-17	NAEP Validity Studies: The Effects of Finite Sampling on State Assessment Sample Requirements	Patricia Dabbs
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown
National Education Longitudinal Study of 1988 (NELS:88)		
95–04	National Education Longitudinal Study of 1988: Second Follow-up Questionnaire Content Areas and Research Issues	Jeffrey Owings
95–05	National Education Longitudinal Study of 1988: Conducting Trend Analyses of NLS-72, HS&B, and NELS:88 Seniors	Jeffrey Owings
95–06	National Education Longitudinal Study of 1988: Conducting Cross-Cohort Comparisons Using HS&B, NAEP, and NELS:88 Academic Transcript Data	Jeffrey Owings
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings

No.	Title	NCES contact
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003-18	Report for Computation of Balanced Repeated Replicate (BRR) Weights for the Third (NELS88:1994) and Fourth (NELS88:2000) Follow-up Surveys	Dennis Carroll
National Household Education Survey (NHES)		
95–12	Rural Education Data User's Guide	Samuel Peng
96–13	Estimation of Response Bias in the NHES:95 Adult Education Survey	Steven Kaufman
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–21	1993 National Household Education Survey (NHES:93) Questionnaires: Screener, School Readiness, and School Safety and Discipline	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
96–29	Undercoverage Bias in Estimates of Characteristics of Adults and 0- to 2-Year-Olds in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
96–30	Comparison of Estimates from the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–02	Telephone Coverage Bias and Recorded Interviews in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–03	1991 and 1995 National Household Education Survey Questionnaires: NHES:91 Screener, NHES:91 Adult Education, NHES:95 Basic Screener, and NHES:95 Adult Education	Kathryn Chandler
97–04	Design, Data Collection, Monitoring, Interview Administration Time, and Data Editing in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–05	Unit and Item Response, Weighting, and Imputation Procedures in the 1993 National Household Education Survey (NHES:93)	Kathryn Chandler
97–06	Unit and Item Response, Weighting, and Imputation Procedures in the 1995 National Household Education Survey (NHES:95)	Kathryn Chandler
97–08	Design, Data Collection, Interview Timing, and Data Editing in the 1995 National Household Education Survey	Kathryn Chandler
97–19	National Household Education Survey of 1995: Adult Education Course Coding Manual	Peter Stowe
97–20	National Household Education Survey of 1995: Adult Education Course Code Merge Files User's Guide	Peter Stowe
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
97–28	Comparison of Estimates in the 1996 National Household Education Survey	Kathryn Chandler
97–34	Comparison of Estimates from the 1993 National Household Education Survey	Kathryn Chandler
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
97–38	Reinterview Results for the Parent and Youth Components of the 1996 National Household Education Survey	Kathryn Chandler
97–39	Undercoverage Bias in Estimates of Characteristics of Households and Adults in the 1996 National Household Education Survey	Kathryn Chandler
97–40	Unit and Item Response Rates, Weighting, and Imputation Procedures in the 1996 National Household Education Survey	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe

BEST COPY AVAILABLE

No.	Title	NCES contact
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
National Longitudinal Study of the High School Class of 1972 (NLS-72)		
95–12	Rural Education Data User's Guide	Samuel Peng
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
National Postsecondary Student Aid Study (NPSAS)		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2002–03	National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), CATI Nonresponse Bias Analysis Report.	Andrew Malizio
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
National Study of Postsecondary Faculty (NSOPF)		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Postsecondary Education Descriptive Analysis Reports (PEDAR)		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Private School Universe Survey (PSS)		
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
96–26	Improving the Coverage of Private Elementary-Secondary Schools	Steven Kaufman
96–27	Intersurvey Consistency in NCES Private School Surveys for 1993–94	Steven Kaufman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Progress in International Reading Literacy Study (PIRLS)		
2003–05	PIRLS-IEA Reading Literacy Framework: Comparative Analysis of the 1991 IEA Reading Study and the Progress in International Reading Literacy Study	Laurence Ogle
2003-10	A Content Comparison of the NAEP and PIRLS Fourth-Grade Reading Assessments	Marilyn Binkley
Recent College Graduates (RCG)		
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Schools and Staffing Survey (SASS)		
94–01	Schools and Staffing Survey (SASS) Papers Presented at Meetings of the American Statistical Association	Dan Kasprzyk
94–02	Generalized Variance Estimate for Schools and Staffing Survey (SASS)	Dan Kasprzyk
94–03	1991 Schools and Staffing Survey (SASS) Reinterview Response Variance Report	Dan Kasprzyk
94–04	The Accuracy of Teachers' Self-reports on their Postsecondary Education: Teacher Transcript Study, Schools and Staffing Survey	Dan Kasprzyk
94–06	Six Papers on Teachers from the 1990–91 Schools and Staffing Survey and Other Related Surveys	Dan Kasprzyk
95–01	Schools and Staffing Survey: 1994 Papers Presented at the 1994 Meeting of the American Statistical Association	Dan Kasprzyk
95–02	QED Estimates of the 1990–91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates	Dan Kasprzyk
95–03	Schools and Staffing Survey: 1990–91 SASS Cross-Questionnaire Analysis	Dan Kasprzyk
95–08	CCD Adjustment to the 1990–91 SASS: A Comparison of Estimates	Dan Kasprzyk

No.	Title	NCES contact
95–09	The Results of the 1993 Teacher List Validation Study (TLVS)	Dan Kasprzyk
95–10	The Results of the 1991–92 Teacher Follow-up Survey (TFS) Reinterview and Extensive Reconciliation	Dan Kasprzyk
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
95–12	Rural Education Data User's Guide	Samuel Peng
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
95–15	Classroom Instructional Processes: A Review of Existing Measurement Approaches and Their Applicability for the Teacher Follow-up Survey	Sharon Bobbitt
95–16	Intersurvey Consistency in NCES Private School Surveys	Steven Kaufman
95–18	An Agenda for Research on Teachers and Schools: Revisiting NCES' Schools and Staffing Survey	Dan Kasprzyk
96–01	Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study	Dan Kasprzyk
96–02	Schools and Staffing Survey (SASS): 1995 Selected papers presented at the 1995 Meeting of the American Statistical Association	Dan Kasprzyk
96–05	Cognitive Research on the Teacher Listing Form for the Schools and Staffing Survey	Dan Kasprzyk
96–06	The Schools and Staffing Survey (SASS) for 1998–99: Design Recommendations to Inform Broad Education Policy	Dan Kasprzyk
96–07	Should SASS Measure Instructional Processes and Teacher Effectiveness?	Dan Kasprzyk
96–09	Making Data Relevant for Policy Discussions: Redesigning the School Administrator Questionnaire for the 1998–99 SASS	Dan Kasprzyk
96–10	1998–99 Schools and Staffing Survey: Issues Related to Survey Depth	Dan Kasprzyk
96–11	Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with comments on School Reform, Governance, and Finance	Dan Kasprzyk
96–12	Predictors of Retention, Transfer, and Attrition of Special and General Education Teachers: Data from the 1989 Teacher Followup Survey	Dan Kasprzyk
96–15	Nested Structures: District-Level Data in the Schools and Staffing Survey	Dan Kasprzyk
96–23	Linking Student Data to SASS: Why, When, How	Dan Kasprzyk
96–24	National Assessments of Teacher Quality	Dan Kasprzyk
96–25	Measures of Inservice Professional Development: Suggested Items for the 1998–1999 Schools and Staffing Survey	Dan Kasprzyk
96–28	Student Learning, Teaching Quality, and Professional Development: Theoretical Linkages, Current Measurement, and Recommendations for Future Data Collection	Mary Rollefson
97–01	Selected Papers on Education Surveys: Papers Presented at the 1996 Meeting of the American Statistical Association	Dan Kasprzyk
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
97–10	Report of Cognitive Research on the Public and Private School Teacher Questionnaires for the Schools and Staffing Survey 1993–94 School Year	Dan Kasprzyk
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–12	Measuring School Reform: Recommendations for Future SASS Data Collection	Mary Rollefson
97–14	Optimal Choice of Periodicities for the Schools and Staffing Survey: Modeling and Analysis	Steven Kaufman
97–18	Improving the Mail Return Rates of SASS Surveys: A Review of the Literature	Steven Kaufman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
97–23	Further Cognitive Research on the Schools and Staffing Survey (SASS) Teacher Listing Form	Dan Kasprzyk
97–41	Selected Papers on the Schools and Staffing Survey: Papers Presented at the 1997 Meeting of the American Statistical Association	Steve Kaufman
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
98–05	SASS Documentation: 1993–94 SASS Student Sampling Problems; Solutions for Determining the Numerators for the SASS Private School (3B) Second-Stage Factors	Steven Kaufman
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
98–12	A Bootstrap Variance Estimator for Systematic PPS Sampling	Steven Kaufman

No.	Title	NCES contact
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
98–14	Variance Estimation of Imputed Survey Data	Steven Kaufman
98–15	Development of a Prototype System for Accessing Linked NCES Data	Steven Kaufman
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Fieldtest Results to Improve Item Construction	Dan Kasprzyk
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
1999–12	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume III: Public-Use Codebook	Kerry Gruber
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Third International Mathematics and Science Study (TIMSS)		
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales

Listing of NCES Working Papers by Subject

No.	Title	NCES contact
Achievement (student) - mathematics		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Adult education		
96–14	The 1995 National Household Education Survey: Reinterview Results for the Adult Education Component	Steven Kaufman
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
98–03	Adult Education in the 1990s: A Report on the 1991 National Household Education Survey	Peter Stowe
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Adult literacy—see Literacy of adults		
American Indian – education		
1999–13	1993–94 Schools and Staffing Survey: Data File User's Manual, Volume IV: Bureau of Indian Affairs (BIA) Restricted-Use Codebook	Kerry Gruber
Assessment/achievement		
95–12	Rural Education Data User's Guide	Samuel Peng
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
97–29	Can State Assessment Data be Used to Reduce State NAEP Sample Sizes?	Larry Ogle
97–30	ACT's NAEP Redesign Project: Assessment Design is the Key to Useful and Stable Assessment Results	Larry Ogle
97–31	NAEP Reconfigured: An Integrated Redesign of the National Assessment of Educational Progress	Larry Ogle
97–32	Innovative Solutions to Intractable Large Scale Assessment (Problem 2: Background Questions)	Larry Ogle
97–37	Optimal Rating Procedures and Methodology for NAEP Open-ended Items	Larry Ogle
97–44	Development of a SASS 1993–94 School-Level Student Achievement Subfile: Using State Assessments and State NAEP, Feasibility Study	Michael Ross
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown

No.	Title	NCES contact
Beginning students in postsecondary education		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
Civic participation		
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Climate of schools		
95–14	Empirical Evaluation of Social, Psychological, & Educational Construct Variables Used in NCES Surveys	Samuel Peng
Cost of education indices		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Course-taking		
95–12	Rural Education Data User's Guide	Samuel Peng
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson
2003–01	Mathematics, Foreign Language, and Science Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
2003–02	English Coursetaking and the NELS:88 Transcript Data	Jeffrey Owings
Crime		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Curriculum		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Customer service		
1999–10	What Users Say About Schools and Staffing Survey Publications	Dan Kasprzyk
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
Data quality		
97–13	Improving Data Quality in NCES: Database-to-Report Process	Susan Ahmed
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown
Data warehouse		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk

BEST COPY AVAILABLE

No.	Title	NCES contact
Design effects		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Dropout rates, high school		
95–07	National Education Longitudinal Study of 1988: Conducting Trend Analyses HS&B and NELS:88 Sophomore Cohort Dropouts	Jeffrey Owings
Early childhood education		
96–20	1991 National Household Education Survey (NHES:91) Questionnaires: Screener, Early Childhood Education, and Adult Education	Kathryn Chandler
96–22	1995 National Household Education Survey (NHES:95) Questionnaires: Screener, Early Childhood Program Participation, and Adult Education	Kathryn Chandler
97–24	Formulating a Design for the ECLS: A Review of Longitudinal Studies	Jerry West
97–36	Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research	Jerry West
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
2001–03	Measures of Socio-Emotional Development in Middle School	Elvira Hausken
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2002-05	Early Childhood Longitudinal Study-Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade	Elvira Hausken
Educational attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Educational research		
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales
Eighth-graders		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Employment		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
Employment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Engineering		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Enrollment – after college		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Faculty – higher education		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler

No.	Title	NCES contact
Fathers – role in education		
2001–02	Measuring Father Involvement in Young Children's Lives: Recommendations for a Fatherhood Module for the ECLS-B	Jerry West
Finance – elementary and secondary schools		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
96–19	Assessment and Analysis of School-Level Expenditures	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Finance – postsecondary		
97–27	Pilot Test of IPEDS Finance Survey	Peter Stowe
2000–14	IPEDS Finance Data Comparisons Under the 1997 Financial Accounting Standards for Private, Not-for-Profit Institutes: A Concept Paper	Peter Stowe
Finance – private schools		
95–17	Estimates of Expenditures for Private K–12 Schools	Stephen Broughman
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Geography		
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
Graduate students		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
Graduates of postsecondary education		
2001–15	Baccalaureate and Beyond Longitudinal Study: 2000/01 Follow-Up Field Test Methodology Report	Andrew G. Malizio
Imputation		
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meeting	Dan Kasprzyk
2001–10	Comparison of Proc Impute and Schafer's Multiple Imputation Software	Sam Peng
2001–16	Imputation of Test Scores in the National Education Longitudinal Study of 1988	Ralph Lee
2001–17	A Study of Imputation Algorithms	Ralph Lee
2001–18	A Study of Variance Estimation Methods	Ralph Lee
Inflation		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
Institution data		
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
Instructional resources and practices		
95–11	Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work	Sharon Bobbitt & John Ralph
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
International comparisons		
97–11	International Comparisons of Inservice Professional Development	Dan Kasprzyk
97–16	International Education Expenditure Comparability Study: Final Report, Volume I	Shelley Burns
97–17	International Education Expenditure Comparability Study: Final Report, Volume II, Quantitative Analysis of Expenditure Comparability	Shelley Burns

No.	Title	NCES contact
2001–01	Cross-National Variation in Educational Preparation for Adulthood: From Early Adolescence to Young Adulthood	Elvira Hausken
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
International comparisons – math and science achievement		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
Libraries		
94–07	Data Comparability and Public Policy: New Interest in Public Library Data Papers Presented at Meetings of the American Statistical Association	Carrol Kindel
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
Limited English Proficiency		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Literacy of adults		
98–17	Developing the National Assessment of Adult Literacy: Recommendations from Stakeholders	Sheida White
1999–09a	1992 National Adult Literacy Survey: An Overview	Alex Sedlacek
1999–09b	1992 National Adult Literacy Survey: Sample Design	Alex Sedlacek
1999–09c	1992 National Adult Literacy Survey: Weighting and Population Estimates	Alex Sedlacek
1999–09d	1992 National Adult Literacy Survey: Development of the Survey Instruments	Alex Sedlacek
1999–09e	1992 National Adult Literacy Survey: Scaling and Proficiency Estimates	Alex Sedlacek
1999–09f	1992 National Adult Literacy Survey: Interpreting the Adult Literacy Scales and Literacy Levels	Alex Sedlacek
1999–09g	1992 National Adult Literacy Survey: Literacy Levels and the Response Probability Convention	Alex Sedlacek
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–05	Secondary Statistical Modeling With the National Assessment of Adult Literacy: Implications for the Design of the Background Questionnaire	Sheida White
2000–06	Using Telephone and Mail Surveys as a Supplement or Alternative to Door-to-Door Surveys in the Assessment of Adult Literacy	Sheida White
2000–07	"How Much Literacy is Enough?" Issues in Defining and Reporting Performance Standards for the National Assessment of Adult Literacy	Sheida White
2000–08	Evaluation of the 1992 NALS Background Survey Questionnaire: An Analysis of Uses with Recommendations for Revisions	Sheida White
2000–09	Demographic Changes and Literacy Development in a Decade	Sheida White
2001–08	Assessing the Lexile Framework: Results of a Panel Meeting	Sheida White
Literacy of adults – international		
97–33	Adult Literacy: An International Perspective	Marilyn Binkley
Mathematics		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
1999–08	Measuring Classroom Instructional Processes: Using Survey and Case Study Field Test Results to Improve Item Construction	Dan Kasprzyk
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	

No.	Title	NCES contact
Parental involvement in education		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
97–25	1996 National Household Education Survey (NHES:96) Questionnaires: Screener/Household and Library, Parent and Family Involvement in Education and Civic Involvement, Youth Civic Involvement, and Adult Civic Involvement	Kathryn Chandler
1999–01	A Birth Cohort Study: Conceptual and Design Considerations and Rationale	Jerry West
2001–06	Papers from the Early Childhood Longitudinal Studies Program: Presented at the 2001 AERA and SRCD Meetings	Jerry West
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
Participation rates		
98–10	Adult Education Participation Decisions and Barriers: Review of Conceptual Frameworks and Empirical Studies	Peter Stowe
Postsecondary education		
1999–11	Data Sources on Lifelong Learning Available from the National Center for Education Statistics	Lisa Hudson
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Postsecondary education – persistence and attainment		
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Postsecondary education – staff		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Principals		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Private schools		
96–16	Strategies for Collecting Finance Data from Private Schools	Stephen Broughman
97–07	The Determinants of Per-Pupil Expenditures in Private Elementary and Secondary Schools: An Exploratory Analysis	Stephen Broughman
97–22	Collection of Private School Finance Data: Development of a Questionnaire	Stephen Broughman
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2000–15	Feasibility Report: School-Level Finance Pretest, Private School Questionnaire	Stephen Broughman
Projections of education statistics		
1999–15	Projected Postsecondary Outcomes of 1992 High School Graduates	Aurora D'Amico
Public school finance		
1999–16	Measuring Resources in Education: From Accounting to the Resource Cost Model Approach	William J. Fowler, Jr.
2000–18	Feasibility Report: School-Level Finance Pretest, Public School District Questionnaire	Stephen Broughman
Public schools		
97–43	Measuring Inflation in Public School Costs	William J. Fowler, Jr.
98–01	Collection of Public School Expenditure Data: Development of a Questionnaire	Stephen Broughman
98–04	Geographic Variations in Public Schools' Costs	William J. Fowler, Jr.
1999–02	Tracking Secondary Use of the Schools and Staffing Survey Data: Preliminary Results	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
2002–02	Locale Codes 1987 - 2000	Frank Johnson

BEST COPY AVAILABLE

No.	Title	NCES contact
Public schools – secondary		
98–09	High School Curriculum Structure: Effects on Coursetaking and Achievement in Mathematics for High School Graduates—An Examination of Data from the National Education Longitudinal Study of 1988	Jeffrey Owings
Reform, educational		
96–03	National Education Longitudinal Study of 1988 (NELS:88) Research Framework and Issues	Jeffrey Owings
Response rates		
98–02	Response Variance in the 1993–94 Schools and Staffing Survey: A Reinterview Report	Steven Kaufman
School districts		
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
School districts, public		
98–07	Decennial Census School District Project Planning Report	Tai Phan
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
School districts, public – demographics of		
96–04	Census Mapping Project/School District Data Book	Tai Phan
Schools		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
2002–02	Locale Codes 1987 – 2000	Frank Johnson
Schools – safety and discipline		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Science		
2000–11	Financial Aid Profile of Graduate Students in Science and Engineering	Aurora D'Amico
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
Software evaluation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
Staff		
97–42	Improving the Measurement of Staffing Resources at the School Level: The Development of Recommendations for NCES for the Schools and Staffing Survey (SASS)	Mary Rollefson
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Staff – higher education institutions		
97–26	Strategies for Improving Accuracy of Postsecondary Faculty Lists	Linda Zimbler
2002–08	A Profile of Part-time Faculty: Fall 1998	Linda Zimbler
Staff – nonprofessional		
2000–13	Non-professional Staff in the Schools and Staffing Survey (SASS) and Common Core of Data (CCD)	Kerry Gruber
State		
1999–03	Evaluation of the 1996–97 Nonfiscal Common Core of Data Surveys Data Collection, Processing, and Editing Cycle	Beth Young
2003-19	NAEP Quality Assurance Checks of the 2002 Reading Assessment Results of Delaware	Janis Brown

No.	Title	NCES contact
Statistical methodology		
97–21	Statistics for Policymakers or Everything You Wanted to Know About Statistics But Thought You Could Never Understand	Susan Ahmed
Statistical standards and methodology		
2001–05	Using TIMSS to Analyze Correlates of Performance Variation in Mathematics	Patrick Gonzales
2002–04	Improving Consistency of Response Categories Across NCES Surveys	Marilyn Seastrom
Students with disabilities		
95–13	Assessing Students with Disabilities and Limited English Proficiency	James Houser
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
Survey methodology		
96–17	National Postsecondary Student Aid Study: 1996 Field Test Methodology Report	Andrew G. Malizio
97–15	Customer Service Survey: Common Core of Data Coordinators	Lee Hoffman
97–35	Design, Data Collection, Interview Administration Time, and Data Editing in the 1996 National Household Education Survey	Kathryn Chandler
98–06	National Education Longitudinal Study of 1988 (NELS:88) Base Year through Second Follow-Up: Final Methodology Report	Ralph Lee
98–11	Beginning Postsecondary Students Longitudinal Study First Follow-up (BPS:96–98) Field Test Report	Aurora D'Amico
98–16	A Feasibility Study of Longitudinal Design for Schools and Staffing Survey	Stephen Broughman
1999–07	Collection of Resource and Expenditure Data on the Schools and Staffing Survey	Stephen Broughman
1999–17	Secondary Use of the Schools and Staffing Survey Data	Susan Wiley
2000–01	1999 National Study of Postsecondary Faculty (NSOPF:99) Field Test Report	Linda Zimbler
2000–02	Coordinating NCES Surveys: Options, Issues, Challenges, and Next Steps	Valena Plisko
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2000–12	Coverage Evaluation of the 1994–95 Public Elementary/Secondary School Universe Survey	Beth Young
2000–17	National Postsecondary Student Aid Study:2000 Field Test Methodology Report	Andrew G. Malizio
2001–04	Beginning Postsecondary Students Longitudinal Study: 1996–2001 (BPS:1996/2001) Field Test Methodology Report	Paula Knepper
2001–07	A Comparison of the National Assessment of Educational Progress (NAEP), the Third International Mathematics and Science Study Repeat (TIMSS-R), and the Programme for International Student Assessment (PISA)	Arnold Goldstein
2001–11	Impact of Selected Background Variables on Students' NAEP Math Performance	Arnold Goldstein
2001–13	The Effects of Accommodations on the Assessment of LEP Students in NAEP	Arnold Goldstein
2001–19	The Measurement of Home Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Graders to Questionnaire Items and Parental Assessment of the Invasiveness of These Items	Arnold Goldstein
2002–01	Legal and Ethical Issues in the Use of Video in Education Research	Patrick Gonzales
2002–02	Locale Codes 1987 - 2000	Frank Johnson
2002–03	National Postsecondary Student Aid Study, 1999–2000 (NPSAS:2000), CATI Nonresponse Bias Analysis Report.	Andrew Malizio
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein
2003-03	Education Longitudinal Study: 2002 (ELS: 2002) Field Test Report	Jeffrey Owings
Teachers		
98–13	Response Variance in the 1994–95 Teacher Follow-up Survey	Steven Kaufman
1999–14	1994–95 Teacher Followup Survey: Data File User's Manual, Restricted-Use Codebook	Kerry Gruber
2000–10	A Research Agenda for the 1999–2000 Schools and Staffing Survey	Dan Kasprzyk
Teachers – instructional practices of		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
2002-06	The Measurement of Instructional Background Indicators: Cognitive Laboratory Investigations of the Responses of Fourth and Eighth Grade Students and Teachers to Questionnaire Items	Arnold Goldstein

No.	Title	NCES contact
Teachers – opinions regarding safety		
98–08	The Redesign of the Schools and Staffing Survey for 1999–2000: A Position Paper	Dan Kasprzyk
Teachers – performance evaluations		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – qualifications of		
1999–04	Measuring Teacher Qualifications	Dan Kasprzyk
Teachers – salaries of		
94–05	Cost-of-Education Differentials Across the States	William J. Fowler, Jr.
Training		
2000–16a	Lifelong Learning NCES Task Force: Final Report Volume I	Lisa Hudson
2000–16b	Lifelong Learning NCES Task Force: Final Report Volume II	Lisa Hudson
Variance estimation		
2000–03	Strengths and Limitations of Using SUDAAN, Stata, and WesVarPC for Computing Variances from NCES Data Sets	Ralph Lee
2000–04	Selected Papers on Education Surveys: Papers Presented at the 1998 and 1999 ASA and 1999 AAPOR Meetings	Dan Kasprzyk
2001–18	A Study of Variance Estimation Methods	Ralph Lee
2003-18	Report for Computation of Balanced Repeated Replicate (BRR) Weights for the Third (NELS88:1994) and Fourth (NELS88:2000) Follow-up Surveys	Dennis Carroll
Violence		
97–09	Status of Data on Crime and Violence in Schools: Final Report	Lee Hoffman
Vocational education		
95–12	Rural Education Data User's Guide	Samuel Peng
1999–05	Procedures Guide for Transcript Studies	Dawn Nelson
1999–06	1998 Revision of the Secondary School Taxonomy	Dawn Nelson

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").